REMARKS

Claims 15, 36, 38-41, 45-46, 56, 61, 63 and 65-69 have been amended; and claims 42, 51 and 52 have been canceled as being redundant in view of the claim amendments made herein. Claims 43-44, 47-49, 50, 53055, 57-60, 62 and 64 remain pending. Upon entry of this amendment, claims 15, 36, 38-41, 43-50, 53-72 will be pending.

Support for the claim amendments and new claims can be found, e.g., at page 17, lines 11-30; page 18, lines 1-16; page 22, lines 13-30, of the specification. No new matter is added.

The claim amendments made herein have been made solely to expedite prosecution of the instant application and should not be construed as an acquiescence to any of the Office's rejections.

Rejection of the Claims Under 35 U.S.C. § 112, First Paragraph

Enablement

The Office has rejected claims 15, 16, 36-60, and 61-69 for allegedly lack of enablement. This rejection is moot as applied to claims 42, 51 and 52, which are now canceled.

In one aspect of the enablement rejection (pages 3-5 of the Office Action), the Office asserted that:

> Claims 15, 16, 36-60 and new claims 61-69 are rejected under 35 U.S.C. 112, first paragraph, because the specification, ... does not reasonably provide enablement for the method which uses any P-selectin LE crystal, which, when the phrase "an amino acid sequence of SEQ ID No: 6, 8 or 9" is broadly interpreted, encompasses **any P-selectin LE that comprises as few as two consecutive amino acids of SEQ ID No: 6, 8 or 9, or those with conservative substitutions thereof, and which further can form in any of the 65 space groups with corresponding unit cell parameters**. Office Action- page 3 (emphasis added)

This aspect of the rejection has been met by amending independent claims 15, 56 and 66 (and claims dependent therefrom) to replace "an" with "the" in the phrase "the amino acid sequence of SEQ ID No: 6, 8 or 9." Thus, these claims, as amended, specify that the P-selectin

LE sequence used in the claimed methods includes the amino acid sequence of SEQ ID NO:6, 8 or 9, or conservative substitutions thereof, instead of "as few as two consecutive amino acids of SEQ ID No: 6, 8 or 9" as alleged by the Office. Claims 15 and 56 have been further amended to be directed to methods of identifying an agent that interacts with P-selectin LE using structural coordinates of the active site of P-selectin having the sequence specified, which are obtained from a crystal of P-selectin LE that has space group $P2_1$ or I222. Thus, these claims encompass methods based on coordinates obtained from crystal of P-selectin belonging to two space groups for which three examples of P-selectin crystals, alone or complexed to two different ligands, are described and characterized in the specification. This aspect of the rejection does not apply to claim 66-71 as these claims do not encompass the use of any P-selectin LE crystal.

In another aspect of the rejection (pages 4-8), the Office concludes that:

> Applicants have met this burden for three P-selectin LE crystals in the specification (described on pp. 31-32); however as stated *supra*, the claims encompass a large number of different protein crystals which have not been described and are by no means trivial to produce. Undue experimentation would be expected in the instant case because even the smallest change in any parameter in crystallizing a protein can have enormous consequences [citing McPherson (Eur. J. Biochem. 1990, 189:1-23)] (Office Action, page 6)

Applicants respectfully traverse this aspect of the enablement rejection as applied to the pending claims, as amended herein.

From the outset, claims 66-71 are directed to methods for identifying agents that interact with P-selectin LE having the amino acid sequence specified (*i.e.*, amino acid sequence of SEQ ID NO:6, 8 or 8, or conservative substitutions thereof), using three dimensional models generated using the full structural coordinates of the active site of P-selectin LE according to Figures 2, 3 or 5, or the selected residues specified, ± a root mean square deviation from the backbone atoms of P-selectin LE of not more than 1.5Å. These claims **do not** require *de novo* crystallization of the P-selectin protein to practice the claimed invention, but instead require manipulation of the structural coordinates specified by the claims to generate three-dimensional models of the active site of P-selectin having the sequences specified to, *e.g.*, perform computer

fitting analysis, or design or select the agent. Thus, the Office's statement regarding the undue experimentation to crystallize a protein is not relevant to these claims.

With respect to the remaining pending claims, the claims have been amended to be directed to methods based on coordinates obtained from crystal of P-selectin belonging to two space groups for which three examples of P-selectin crystals, alone or complexed to two different ligands, are described and characterized in the specification. More specifically, at least three crystallization conditions for generating P-selectin LE crystals having space group $P2_1$ or I222 in uncomplexed form and complexed with two different ligands, *i.e.*, SLe^x and PSGL-1 are described in the instant Examples. The structural coordinates for each of these different forms of P-selectin are disclosed in the specification in Figures 2, 3 and 5, respectively. In addition, the structural coordinates of E-selectin LE, another member of the selectin family, complexed with SLe^x was provided in Figure 4. The LE-domains of P- and E-selectin share 62% identity at the amino acid level. Therefore, the present applications describes crystallization conditions and structural coordinates for P-selectin LE in three different forms, as well as the coordinates of a related selectin family member sharing 62% identity at the amino acid level complexed SLe^x. Optimized conditions for generating crystals of the aforesaid P-selectin variants are described starting, *e.g.*, at page 34, line 17 through page 36, line 30, of the specification.

Applicants submit that once crystallization parameters are established and optimized (as it is the case in the instant application), one of ordinary skill in the art would have been able to generate *de novo* crystals of P-selectin LE in uncomplexed or complexed form within the space groups specified, without undue experimentation. This conclusion is consistent with the McPherson reference cited by the Office when it provides:

> Macromolecular crystallization is, thus, a matter of searching, as systematically as possible, the ranges of the individual parameters that impact upon crystal formation, finding a set or multiple set of these factors that yield some kind of crystal, and then optimizing the variable sets to obtain the best possible crystals for X-ray analysis. (McPherson, A. (1990) *Eur. J. Biochem.* 189: 1-23)

The present disclosure details the experimental conditions with the parameters that impact P-selectin LE crystal formation and optimization. Once that information is provided, one

of ordinary skill in the art would have been able to generate crystals of P-selectin LE with the properties specified in the claims by practicing routine experimentation.

The Office further states that claims 61-65 that recite the specific space group and unit cell parameters and ligand binding are "still deemed beyond the scope of enablement because the independent claims are unlimited as to the amino acid sequences used to make said crystals, e.g., the claims encompass any P selectin LE protein comprising as few as two contiguous amino acids of SEQ ID NO: 6, 8, or 9 and further comprises those proteins with conservative substitutions." It is noted that the independent claims have been amended herein to encompass P-selectin sequences having the sequence specified and conservative substitutions thereof. Claims 61, 63 and 65 have been further amended to specify the particular sequences and space group with unit cell parameters recited by the claim. Thus, reconsideration and withdrawal of this aspect of the rejection is respectfully requested in view of the claim amendments made herein.

On pages, 19-24 of the Office Action, the Office re-articulates the position that "even changes as few as one or two amino acids can have large consequences." The Office further states that "[i]t is in no way apparent or limiting that the structural coordinates from the crystal are one and the same as the structural coordinates of said relative structural coordinates of said Figures 2, 3 or 5." (Office Action pages 19-20).

This aspect of the rejection is met in part by the amendments to claims 15 and 56, and is traversed in part. Claims 15 and 56 have been amended to more clearly recite that the relative structural coordinates used in the claimed methods are obtained from the P-selectin crystals provided having space group $P2_1$ or I222.

The Office acknowledged Applicants's comments in the previous response on the high level of skill in the art in making changes (e.g., conservative substitutions) for proteins that are "soluble and require biological activity." (Office Action at page 20). However, the Office maintains that as few as one or two amino acids can have "large consequences" to the reproducibility of protein crystals. In response, Applicants submit that it was known in the art, at the time the instant application was filed, that protein variants with known crystallization

parameters were likely to readily crystallize with similar crystal structures as long the variations introduced did not markedly affect intermolecular crystal contacts or amino acid residues important for protein stability (*i.e.*, within the hydrophobic core). *See* Itoh, S. I. and M. A. Navia (1995) *Protein Science*, (4), 2261-2268 (copy submitted herewith as Exhibit A). Even mutations that had an effect in altering protein stability were found to crystallize with similar crystallization parameters as the native protein, emphasizing that well-folded proteins can exhibit crystallization properties similar to the non-mutated counterparts. *See* Sauer, U. H., S. Dao-Pin, and B. W. Matthews (1992) *Journal of Biological Chemistry* (267) 2393-2399 (copy submitted herewith as Exhibit B). Therefore, in view of the teachings in the specification describing successful crystallization conditions for at least three variants of P-selectin LE and E-selectin, one of ordinary skilled in the art would have been able to generate the conservatively substituted variant of P-selectin and produce crystals of these variants, without undue experimentation.

In view of the foregoing, reconsideration and withdrawal of this rejection is respectfully requested.

Written Description

On pages 8-12 of the Office Action, the Office has rejected claims 15, 16, 36-60 and 61-69 for alleged lack of written description. This rejection is moot as applied to claims 42, 51 and 52, which are now canceled.

In one aspect of this rejection, the Office states that the claims are:

> [I]ntrinsically drawn to a large number of species of P-selectin crystals containing a considerable number of different P-selectin proteins and thus the claims possess a large genus of widely variant crystals of both P-selectin proteins used to make the crystals, as well as a large genus of widely variant crystal forms themselves (e.g., any of the 65 space groups). Also, as noted above, the crystals can have any ligand bound to P-selectin LE. However, the specification only adequately describes three representative species in terms of both structure and function which belong to this genus.

As noted above, this aspect of the rejection has been met by amending independent claims 15, 56 and 66 (and claims dependent therefrom) to replace "an" with "the" in the phrase

"the amino acid sequence of SEQ ID No: 6, 8 or 9." Thus, these claims, as amended, specify that the P-selectin LE sequence used in the claimed methods includes the amino acid sequence of SEQ ID NO:6, 8 or 9, or conservative substitutions thereof, instead of "as few as two consecutive amino acids of SEQ ID No: 6, 8 or 9 as alleged by the Office. Claims 15 and 56 have been further amended to be directed to methods of identifying an agent that interacts with P-selectin LE using structural coordinates of the active site of P-selectin having the sequence specified, which are obtained from a crystal of P-selectin LE that has space group $P2_1$ or I222. Thus, these claims encompass methods based on coordinates obtained from crystal of P-selectin belonging to two space groups (instead of "any of the 65 space groups") for which three examples of P-selectin crystals, alone or complexed to two different ligands, are described and characterized in the specification.

As stated above in response to the enablement rejection, claim 66 does not require the use of a P-selectin LE crystal, thus Applicants request reconsideration of the Office's position quoted above as applied to claim 66 and claims dependent therefrom.

Related to this rejection, the Office has taken the position that the crystals can have "any ligand bound to P-selectin LE. However, the specification only adequately describes three representative species." Applicants respectfully traverse this aspect of the rejection. The claims recite structural coordinates for P-selectin crystals in uncomplexed form (which theoretically can have "any ligand bound to P-selectin LE") and complexed with two physiological ligands of P-selectin LE, *i.e.*, PSGL and SLe^x. The structural coordinates of three species of human P-selectin in uncomplexed form and complexed to SLe^x and PSGL-1 were identified as set forth in Figures 2, 3 or 5, respectively, of the application. In addition, the structural coordinates of E-selectin LE, another member of the selectin family, complexed with SLe^x was provided in Figure 4. The LE-domains of P- and E-selectin share 62% identity at the amino acid level. Therefore, the present applications describes the structural coordinates for P-selectin LE in three different forms, as well as a related selectin family member sharing 62% identity at the amino acid level. Applicants submit that the breadth of the P-selectin LE genus specified by the method claims is circumscribed to the recited relative structural coordinates, which are obtained from the active

site of the P-selectin crystals having space group $P2_1$ or I222 recited by the claims. Given the sequence identity and structural limitations encompassed by the methods claims as amended herein, the claims provide sufficient characteristics in common to define the genus of crystals recited by the claimed methods.

Lastly, claims 61-65 that recite the specific space group and unit cell parameters and ligand binding are "still deemed to not possess written description because the independent claims are unlimited as to the amino acid sequences used to make said crystals, particularly in view of the broad by (*sic.*) reasonable interpretation of the claims, which encompass P-selectin LE polypeptides have as few as two consecutive amino acids of SEQ ID NO: 6, 8, or 9, and those which also encompass conservative substitutions there of." It is noted that the independent claims have been amended herein to encompass P-selectin sequences having the sequence specified and conservative substitutions thereof. Claims 61, 63 and 65 have been further amended to specify the particular sequences and space group with unit cell parameters recited by the claim. Claims 62 and 64, which depend from claims 15 and 56, specify the ligand bound by the P-selectin LE complex of the claim. Thus, reconsideration and withdrawal of this aspect of the rejection is respectfully requested in view of the claim amendments made herein.

Given the knowledge in the art and the teachings of the specification, a skilled practitioner at the filing date would have recognized that Applicants were in possession of, and had adequately described, the P-selectin LE genus recited in the pending claims to satisfy the standard set forth in the MPEP (*see, e.g.*, MPEP §2163(II)(A)(3)(ii)). Accordingly, Applicants submit that the claims, as presently pending, fully satisfy the written description requirement and reconsideration of this rejection is respectfully requested.

Rejection of Claims 66-69 under 35 U.S.C. § 103(a)

On pages 13-17 of the Office Action, the Office has rejected claims 66-69 under 35 USC §103(a) as being unpatentable over Revelle *et al.* (1996) JBC 271 (27): 16160-16170 in view of Morris *et al.* (1996) *J. of Computer-Aided Molecular Design* 10: 293-304 in view of *In re Gulack*

217 USPQ 401 (Fed. Cir. 1983) and *In re Ngai* USPQ2d 1862 (Fed. Cir. 2004). According to the Office:

> All claim limitations concerning the machine readable data comprising structure coordinate data of Figures 2, 3 and 5 are given no patentable weight as it is considered to be non-functional descriptive material. As such, the instant claims are considered to be limited to a method of using a known computer program to identify agents that interact with P-selectin by inputing the three-dimensional structural coordinates into said computer program, and analyzing the output by visual/mental interpretation....
>
> Therefore it would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains to utilize the program AutoDock (developed by Morris et al.) which is specifically used for identifying agents/ligands that interact with macromolecular structures and to use said program with any three-dimensional structural coordinates, including those of the instant claims/invention. (Office Action at pages 15 and 17)

The primary reference by Revelle *et al.* is simply a review of the effects of varions mutations in E- and P-selectin in determining carbohydrate binding specifity. There is no teaching or suggestion in this reference regarding three dimensional structures or models of the active site of P-selectin LE, let alone a method for identifying P-selectin LE using particular structural coordinates of P-selectin LE, alone or complexed with its physiological ligands. The secondary reference of Morris *et al.* fails to make up for the deficiencies in Revelle *et al.* as it simply discloses a general description of software programs for designing and determining potential ligand-protein interactions. The Office's rejection rests on the assumption that the "structure coordinate data of Figures 2, 3 and 5 are given no patentable weight as it is considered to be non-functional descriptive material."

Applicants respectfully traverse this rejection as applied to claims 66-69.

Claim 66 and claims dependent thereon are directed to a method of identifying an agent that interacts with P-selectin LE having the relative structural coordinates of the active site of P-selectin LE binding site specified by the claimed methods; generating a three dimensional model based on the relative structural coordinates; and evaluating the fit between the three dimensional model of the active site and the candidate agent, *e.g.*, by computer fitting analysis. Applicants submit that the structural coordinates of Figures 2, 3 and 5 evaluated by the claimed methods

impart functionality by changing the processing steps of the computer program, changing the structural coordinates of the P- selectin LE binding site and the candidate agent, which ultimately imposes a change in the screening and/or design process that leads to obtaining an agent that interacts with P- selectin LE.

When the candidate agent is positioned in the active site of P-selectin LE, the particular structural coordinates of the P-selectin LE site recited in the claims provide a specific spatial relationship and energy surface between the binding site and the candidate agent. During the docking process, the orientation of the candidate agent is constantly adjusted in the binding site by interactive real-time energy calculations between the binding site and the candidate agent. The energy calculations provide feedback to the docking program and dictate how the computer program functions to find an energetically favorable conformation of the candidate agent. If the interaction between the binding site and the candidate agent moves uphill in energy, this feedback will dictate the computer program to resist the motion. If the interaction between the binding site and the candidate agent is favorable, the feedback with dictate the computer program to encourage the motion (*See* N. Claude-Cohen *et al.* (1990) *J. of Med. Chemistry* 33(3):883-894, submitted herewith as Exhibit C). Thus, the structural coordinates of the P- selectin LE binding site recited in the claims dictate how the computer program functions.

Furthermore, the structural coordinates of the binding site are not merely used for comparison of the structural coordinates of the candidate agent. As specified in the claims, the structural coordinates of the candidate agent are in fact changed by the structural coordinates of the P- selectin LE binding site during the docking process. As the orientation of the candidate agent is adjusted, the three-dimensional structural coordinates of the candidate agent are changed.

Therefore, the structural coordinates of the P- selectin LE binding site specified by the claimed methods impart functionality by changing the processing steps of the computer program, changing structural coordinates of the candidate agent, which ultimately imposes a change in the screening and/or design process that leads to obtaining an agent that interacts with P- selectin LE. Such structural information is <u>not</u> non-functional descriptive material, as alleged by the

Office, as it imparts a series of concrete steps having a functional relationship between matter and substrate. *See Gulack*, 703 F.2d at 1387.

Accordingly, reconsideration and withdrawal of the present rejection are respectfully requested.

Conclusion

Reconsideration and withdrawal of the present rejections are respectfully requested. If the Office would like to contact the undersigned attorney, she may do so by calling (617) 368-2131.

Enclosed is a check for the Petition for Extension of Time fee. Please apply any other charges or credits to deposit account 06-1050, referencing attorney docket number 16163-004001.

Respectfully submitted,

Date: December 14, 2007

Diana Collazo
Reg. No. 46,635

Fish & Richardson P.C.
225 Franklin Street
Boston, MA 02110
Telephone: (617) 542-5070
Facsimile: (617) 542-8906

21732163.doc

Appendix 17

Protein Science (1995), *4*:2261–2268. Cambridge University Press. Printed in the USA.

Structure comparison of native and mutant human recombinant FKBP12 complexes with the immunosuppressant drug FK506 (tacrolimus)

SUSUMU ITOH[1] AND MANUEL A. NAVIA
Vertex Pharmaceuticals Incorporated, Cambridge, Massachusetts 02139-4211

(RECEIVED June 13, 1995; ACCEPTED August 9, 1995)

Abstract

The consequences of site-directed mutagenesis experiments are often anticipated by empirical rules regarding the expected effects of a given amino acid substitution. Here, we examine the effects of "conservative" and "nonconservative" substitutions on the X-ray crystal structures of human recombinant FKBP12 mutants in complex with the immunosuppressant drug FK506 (tacrolimus). R42K and R42I mutant complexes show 110-fold and 180-fold decreased calcineurin (CN) inhibition, respectively, versus the native complex, yet retain full peptidyl prolyl isomerase (PPIase) activity, FK506 binding, and FK506-mediated PPIase inhibition. Interestingly, the structure of the R42I mutant complex is better conserved than that of the R42K mutant complex when compared to the native complex structure, within both the FKBP12 protein and FK506 ligand regions of the complexes, and with respect to temperature factors and RMS coordinate differences. This is due to compensatory interactions mediated by two newly ordered water molecules in the R42I complex structure, molecules that act as surrogates for the missing arginine guanidino nitrogens of R42. The absence of such surrogate solvent interactions in the R42K complex leads to some disorder in the so-called "40s loop" that encompasses the substituent. One rationalization proposed for the observed loss in CN inhibition in these R42 mutant complexes invokes indirect effects leading to a misorientation of FKBP12 and FK506 structural elements that normally interact with calcineurin. Our results with the structure of the R42I complex in particular suggest that the observed loss of CN inhibition might also be explained by the loss of a specific R42-mediated interaction with CN that cannot be mimicked effectively by the solvent molecules that otherwise stabilize the conformation of the 40s loop in that structure.

Keywords: calcineurin; immunophilins; site-directed mutagenesis; structure-based drug design; X-ray crystallography

FK506 (United States Adopted Names Council of the American Medical Association, Chicago, Illinois [USAN], *tacrolimus*) is a natural product screening lead (Kino et al., 1987) now approved for therapeutic use as an immunosuppressant in Japan, the USA, Germany, and other countries. FK506, in complex with its 12-kDa M_r binding protein FKBP12, exerts its immunosuppressive effects through the inhibition of calcineurin (CN), an intracellular Ca^{+2}-calmodulin-dependent phosphatase (Klee & Cohen, 1988). CN inhibition, in turn, interrupts the induction of IL-2 and other T-cell activation events (Friedman & Weissman, 1991; Liu et al., 1991). A homologous natural product, rapamycin (AY-22,989; USAN, *sirolimus*), which was initially discovered as an antifungal agent (Sehgal et al., 1975), can antagonize the CN inhibitory activity of FK506 (Bierer et al., 1990a; Dumont et al., 1990b), even though it is itself an immunosuppressant by a different mechanism (Bierer et al., 1990a; Dumont et al., 1990a). These differences in CN inhibitory activity between the agonist FK506 and the antagonist rapamycin in their complexes with FKBP12 (Table 1), were first explained within the framework of an elegant model (Schreiber, 1991) that focused attention on the corresponding differences in chemical structure between the two ligands in their so-called "effector domain" region (see, e.g., Fig. 1). X-ray crystallographic studies have provided support for this model (van Duyne et al., 1991a, 1991b, 1993; Becker et al., 1993; Rotonda et al., 1993; Connelly et al., 1994; Wilson et al., 1995), by showing that the effector

Reprint requests to: Manuel A. Navia, Vertex Pharmaceuticals Incorporated, 40 Allston Street, Cambridge, Massachusetts 02139-4211; e-mail: navia@vpharm.com.

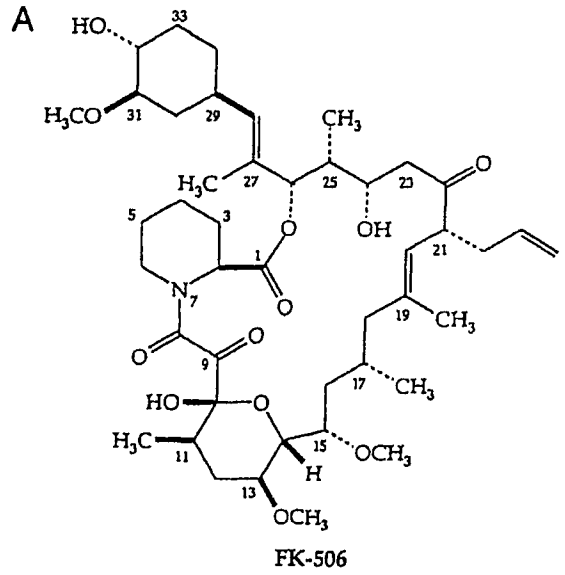
[1] Visiting from: Chugai Pharmaceuticals Co. Ltd., 1-135 Komakado, Gotemba Shizuoka 412, Japan.

Abbreviations: CN, calcineurin; PPIase, peptidyl prolyl isomerase; FK506, USAN tacrolimus; FKBP12, 12-kDa M_r FK506 binding protein; FKBP13, 13-kDa M_r FK506 binding protein; USAN, United States Adopted Names Council of the American Medical Association, Chicago, Illinois.

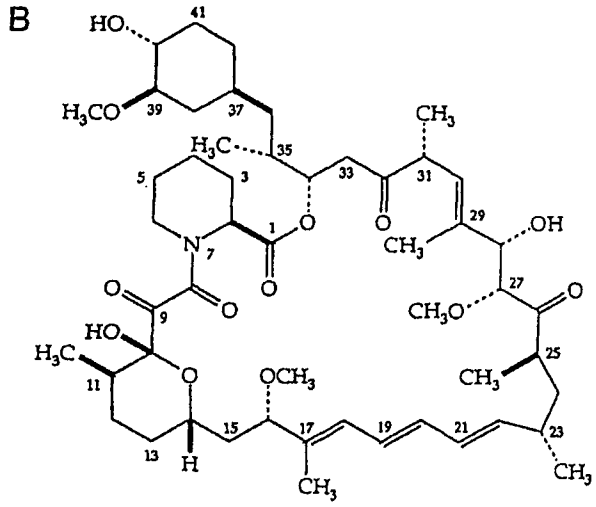
Table 1. *Biochemical properties of the native and mutant complexes*[a]

Mutant	Calcineurin K_i (nM)	Reduction vs. native complex	FK506 K_i (nM)	Specific PPIase activity ($s^{-1}\ \mu M^{-1}$)
Native[b]	5.5 (1.8)	—	0.6 (0.2)	4.3 (0.4)
R42K[b]	590 (200)	107.3	0.6 (0.2)	3.8 (0.3)
R42I[b]	970 (150)	176.4	0.1 (0.1)	2.5 (0.3)
R42Q[c]	325 (150)	59.1	4.3 (2.0)	3.0 (0.3)
Native[d]	7.9 (3.0)	—	0.4 (0.2)	2.2 (0.2)
R42Q[d]	850 (250)	107.6	1.7 (0.6)	1.3 (0.3)
R42A[d]	280 (80)	35.4	0.2 (0.1)	1.1 (0.2)
Chimera[d,e]	19 (2)	2.4	0.4 (0.2)	0.57 (0.05)

[a] Summary of published biochemical data for native and mutant FKBP12 proteins and their complexes with FK506 and calcineurin. The inhibition constant for calcineurin by native and mutant FKBP12 complexes with FK506 is given, along with the FKBP12 inhibition constants versus FK506 and the PPIase specific activity of the various FKBP12s versus a synthetic substrate.
[b] Aldape et al. (1992).
[c] Futer et al. (1995).
[d] Yang et al. (1993).
[e] Substitute FKBP12 residues 40-44 (-RDRNK-) with the corresponding residues (-LPQNQ-) from FKBP13.



FK-506



Rapamycin

Fig. 1. A: Chemical structure of FK506 (USAN, tacrolimus). The effector domain of FK506 (Schreiber, 1991) corresponds to those portions of the ligand (C18–C23 and substituents in the macrocycle, and C26–C34 in the cyclohexyl ring) that have been shown crystallographically to protrude from the surface of its complex with FKBP12; structural elements in common between rapamycin and FK506 have been shown crystallographically to bind the PPIase active site of FKBP12 in the same manner (van Duyne et al., 1991a, 1991b, 1993; Becker et al., 1993; Rotonda et al., 1993; Armistead et al., 1995; Itoh et al, 1995; Wilson et al., 1995). **B:** Chemical structure of rapamycin (AY-22989, USAN, sirolimus). The rapamycin effector domain corresponds to atoms C15–C29 in the macrocycle and C36–C42 in the cyclohexyl ring.

domains of FK506 and rapamycin do indeed protrude from the surface of their respective protein–ligand complexes (Fig. 2A) with distinct conformations that might be compatible with CN binding and inhibition for the one ligand, but not for the other. In turn, those chemical structure elements shared by the two ligands (Fig. 1) were shown in those studies to constitute FKBP12 "binding domains," allowing a rationalization of the reciprocal antagonism between the two ligands in terms of their competition for a common FKBP12 binding site.

The effector domain model has retained broad acceptance as a first approximation to the complicated problem of immunosuppressive drug design, in part because of its consistency with the observed loss of CN inhibitory activity that follows even minor variation in the chemical structure of FK506 (Goulet et al., 1994). By limiting the role of the FKBP12 protein to that of a presenter of ligand functionality to CN (Schreiber, 1991; Rosen & Schreiber, 1992; Schreiber et al., 1993), the model reduces the scope of the drug design problem to one of simple mimicry of the conformation of the FK506 effector domain that protrudes from the surface of the native FKBP12–FK506 complex. Unfortunately, these efforts have yet to produce a linear or macrocyclic drug lead—let alone a clinical candidate—that exceeds the potency of FK506 (Itoh et al., 1995); ligands predicted on the basis of this model have all turned out to be antagonists of FK506 (see, e.g., Bierer et al., 1990b; Somers et al., 1991; Armistead et al., 1995).

Experimental evidence for a more complicated interaction between CN and the FKBP12–FK506 complex first emerged from a systematic examination of the biochemical properties of site-directed mutants of charged residues on the surface of FKBP12 (Aldape et al., 1992). The critical involvement of the "40s loop" and "80s loop" regions of the protein (as defined in Fig. 2B) was established for this interaction by these and subsequent studies (Yang et al., 1993; Futer et al., 1995), leading to a generalization of the effector domain model in the direction of a composite "effector surface" of both protein and ligand structural elements. Elsewhere, we have explored the structural consequences that follow substitutions in the 80s loop region of FKBP12 (Itoh et al., 1995) and have identified composite features on the FKBP12-

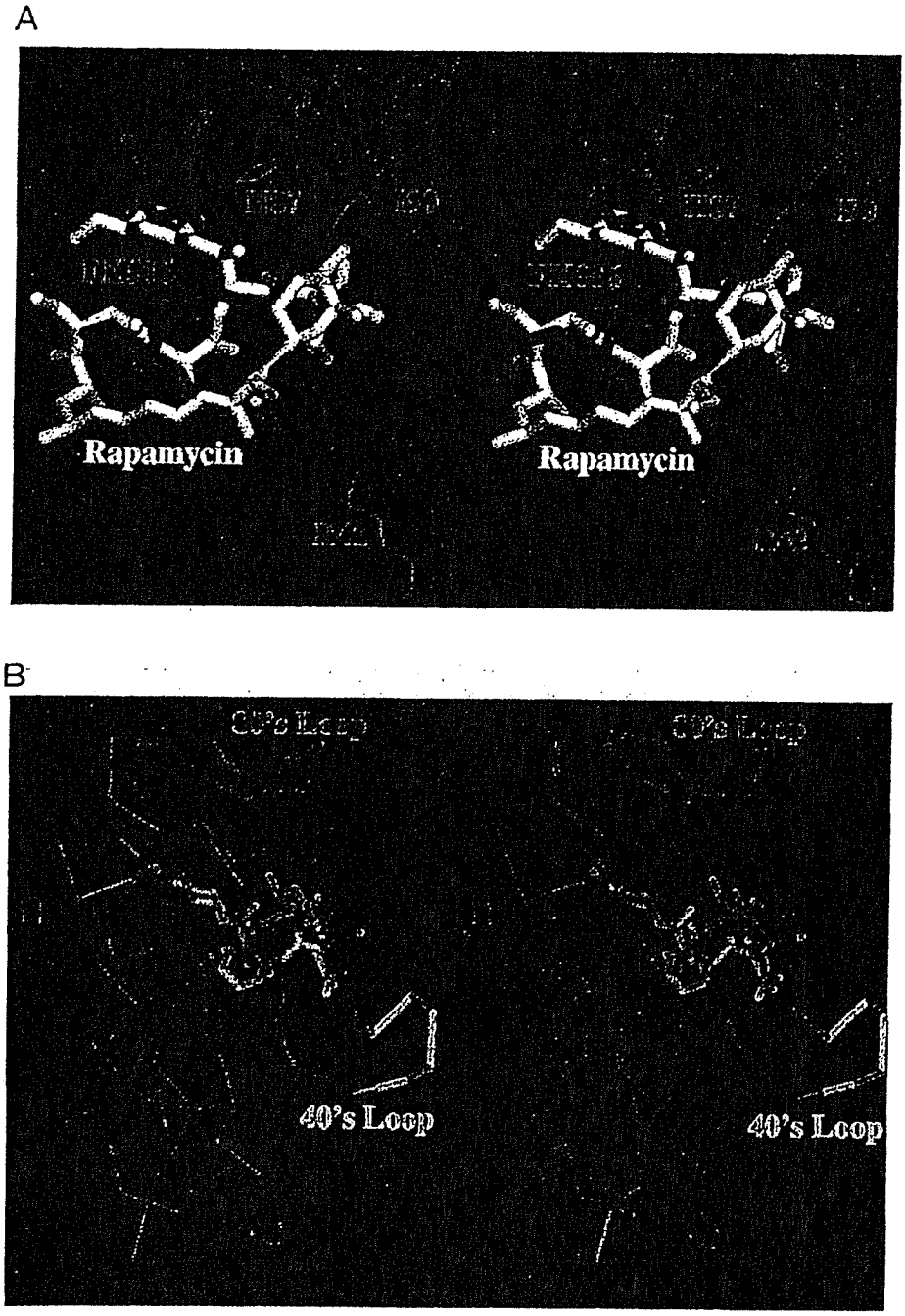


Fig. 2. **A:** Comparison of the structures of the FKBP12 complexes with FK506 (in red) and with rapamycin (in white). A dot representation of the surface of native FKBP12 (Wilson et al., 1995) is shown in blue and is representative of the surface of the FKBP12 complex structures with FK506 and rapamycin. **B:** Conformation of the backbone Cα of FKBP12 in its complex with FK506 is shown in blue. The 40s loop and 80s loop regions of the FKBP12 protein are specifically identified in yellow and red, respectively. The 40s loop is made up of residues 40–44 in FKBP12, which form a bulge on the third β-strand of the protein, as defined by van Duyne et al. (1991a). The 80s loop includes residues 84–91 on the edge of the fifth β-strand of the structure.

FK506 effector surface that might be candidate CN recognition and binding elements (Wilson et al., 1995). In this paper we focus on structural elements in the 40s loop region of FKBP12 and, in particular, on substitutions at residue 42 of the protein, which demonstrate profound (Table 1) but complicated effects on the CN inhibitory potency of the corresponding mutant complexes with FK506. These mutant data have led to proposed mechanisms of action, which are distinctly different in their character and consequences, that need to be resolved (Clardy, 1995).

Results

Crystallization, data collection, and refinement statistics for the native and mutant complex structures reported here are given in Table 2, along with RMS differences in conformation and mobility versus the native complex structure; biochemical data are summarized from the existing literature (Aldape et al., 1992; Yang et al., 1993; Futer et al., 1995) in Table 1. In all the complexes studied, the FKBP12 fold (Fig. 2B) that was seen in the

Table 2. *Summary of the wild-type and mutant complex structure analyses*[a]

	Wild type	R42K	R42I
Area detector used	Siemens	Rigaku	Rigaku
$P4_32_12$ unit cell; a, c (Å)	58.39, 55.76	58.31, 55.93	58.25, 55.98
Resolution (Å)	6.0–1.5	6.0–1.5	6.0–1.6
No. observations	76,344	43,803	44,188
% Reflection ($I > 2\sigma$)	81.4	88.1	89.7
R-merge (%)	3.99	5.88	3.15
R-factor (%)	16.6	18.7	16.8
No. water molecules	85	83	87
RMS bond length error (Å)	0.016	0.016	0.018
RMS bond angle error (deg)	2.74	2.85	2.93
Avg. FK506 B-factors (Å^2)	12.3	15.3	11.0
FK506 RMS diff vs. wt (Å)	–	0.139	0.123
FKBP12 RMS diff vs. wt (Å)	–	0.146	0.147

[a] Native and mutant FKBP12 complexes all share the native crystal form first reported by van Duyne et al. (1991a).

native complex structure (van Duyne et al., 1991a) is strongly conserved, consistent with the observed retention of PPIase and FK506 binding activity (Table 1). In this study, a considerable effort was made to crystallize all the complexes reported in a common crystal form (Table 2), in order to facilitate a direct comparison between structures. This crystal form turned out to be that of the native FKBP12-FK506 complex (van Duyne et al., 1991a), as a consequence of seeding mutant complex crystallization experiments with microcrystals of the native complex and subsequently using crystals from those solutions as macroseeds leading to data quality mutant complex crystals.

In the native complex structure (red coordinates in Fig. 3A, Kinemage 1), the two guanidino nitrogens of R42 are seen to stabilize the "40s loop" of FKBP12 through their participation in a bridging network of noncovalent interactions between residues D37 and K44. In the R42I mutant complex structure (yellow coordinates in Fig. 3A, Kinemage 1), two tightly bound water molecules (yellow in Fig. 3A, Kinemage 1) substitute for the missing

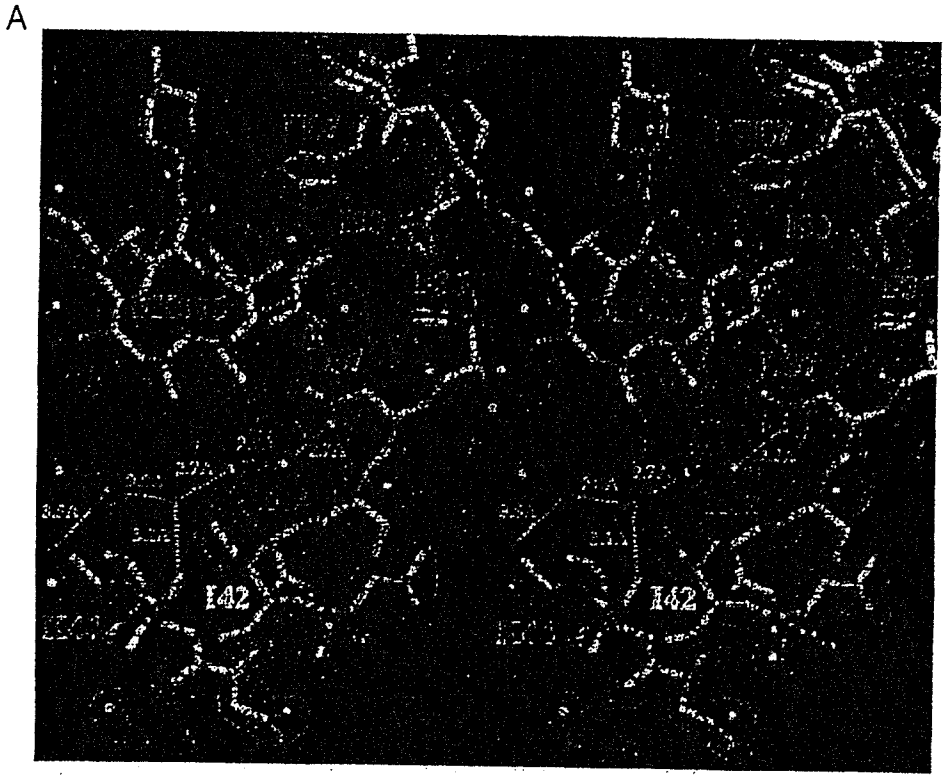


Fig. 3. A: Comparison of the structures of native (red) and R42I mutant (yellow) FKBP12 complexes with FK506. Refined coordinates (Table 2) of both complex structures are superimposed in the figure, together with the $2|F_o| - |F_c|$ electron density (in blue) of the R42I mutant complex structure, contoured at 1σ above background. Part of the FK506 binding domain is shown, as well as residues surrounding R42 and H87 in the protein. Water molecules bridging residues D37 and K44 are shown in green and yellow, corresponding to the R42I mutant complex structure. Dashed lines indicate bond distances between the water molecules. Our observations are inconsistent with the suggestion of Yang et al. (1993) that R42 mutants exert their effects indirectly, by reorienting nearby regions of the complex structure. Nor does the conformation of the 40s loop change as drastically as that of FKBP13 (Schultz et al., 1994) as a consequence of these substitutions. B: Dot-surface representation of the R42I mutant complex structure in the vicinity of FK506. Two water molecules (shown in green) in the R42I mutant complex structure fit readily into the gap created by the R42I substitution and act as surrogates for the missing guanidino nitrogens of R42 in bridging residues D37–K44. This interaction helps preserve the native conformation in the mutant complex and maintains PPIase and FK506 binding activity. Nonetheless, CN inhibition is drastically reduced (Table 1), suggesting a specific protein–protein interaction in the native complex that the water surrogates would be unable to mimic. C: Structure of the R42K mutant complex (in green) compared to that of the native complex (in red). As above, the $2|F_o| - |F_c|$ electron density corresponding to the R42K mutant complex, contoured at 1σ above background, is presented in blue. (*Continues on facing page.*)

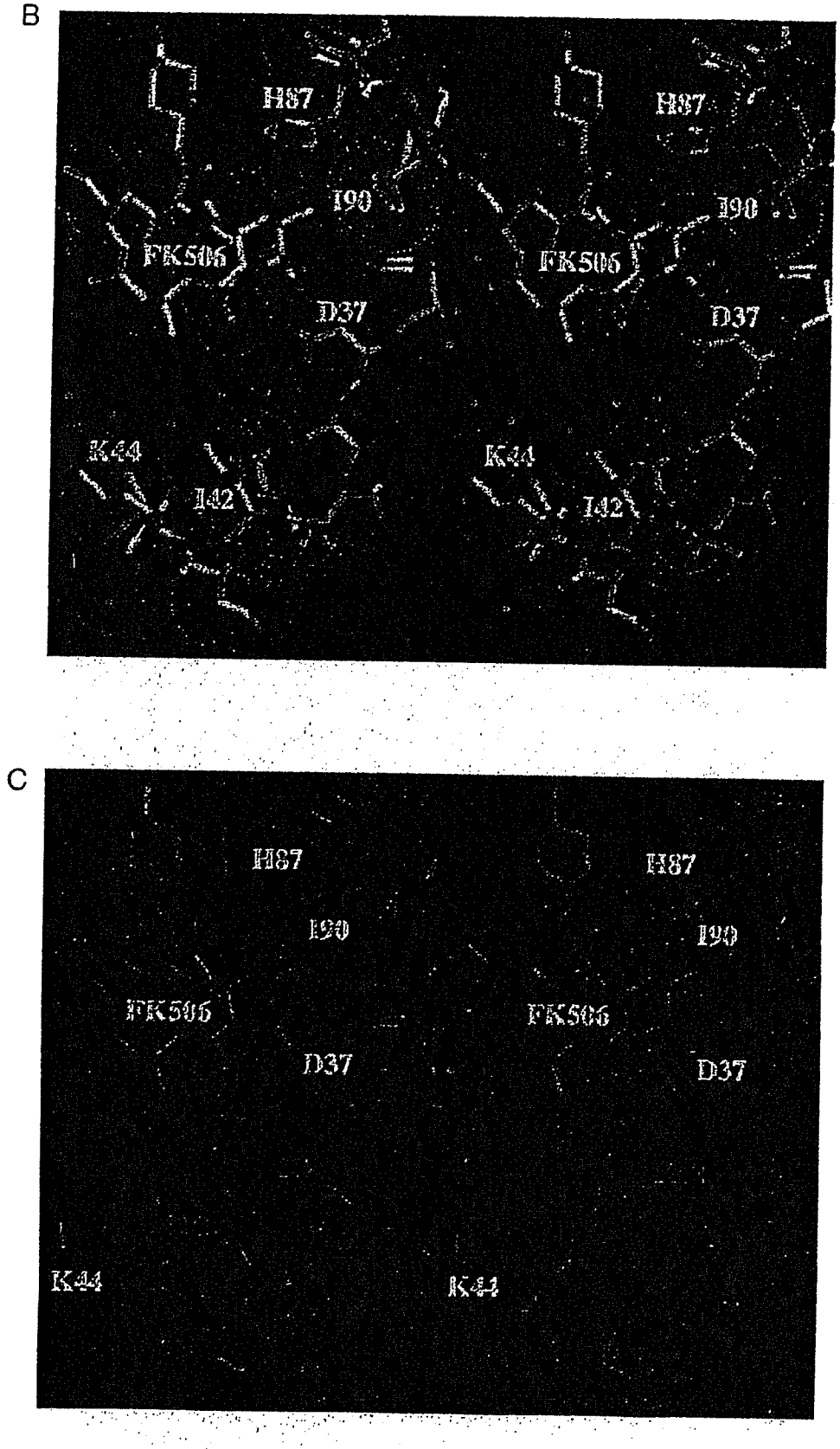


Fig. 3. *Continued.*

arginine guanidino nitrogens of R42, and are well accommodated (Fig. 3B, Kinemage 1) in the gap created by the smaller isoleucine substituent; a third water molecule (green in Fig. 3A, Kinemage 1) is apparently common to all the FKBP12 complex structures. As guanidino nitrogen surrogates, the two water molecules help to preserve the native complex 40s loop conformation in the R42I mutant complex, with an RMS deviation of 0.147 Å between the structures (Table 2). This interaction resembles one seen in the structure of a T157G mutant of T4 lysozyme (Alber et al., 1987; Matthews, 1993), where an ordered water molecule, acting as a surrogate for the missing T157 hydroxyl, preserves the pattern of stabilizing hydrogen bonding interactions seen in the native protein. It is interesting that, in spite of the high degree of conservation seen in the structure of the R42I mutant complex, CN inhibition is nonetheless reduced by ~180-fold (Table 1).

In the structure of the FK506 complex with the more conservative R42K mutant protein, the single ε-amino nitrogen of the lysine side chain is unable to substitute for both of the guanidino nitrogens of arginine (red in Fig. 3C, Kinemage 1); nor does the substitution leave enough space for an additional water molecule to insert itself as a surrogate, as shown in Figure 3B (Kinemage 1) for the R42I mutant complex. The resulting destabilization of the 40s loop is reflected in the higher temperature factors observed in that region of the protein (Fig. 4), even though the conformation of mutant protein complex still closely resembles that of the native (Fig. 3C, Kinemage 1), and FK506 binding and PPIase activity are preserved (Table 1).

Discussion

Two working models have emerged to rationalize the profound though complicated effects on CN inhibition that accompany substitutions in and around residue 42 of FKBP12. The simpler of these is made evident in the R42 single-site mutant complexes that were first characterized biochemically by Aldape et al. (1992), whose crystal structures are described here. In that model, the observed loss of CN-inhibitory activity can be immediately explained in terms of a direct and localized perturbation, by the substituted residue, of the effector surface presented to CN by the corresponding FKBP12–FK506 complex. The Merck group (Becker et al., 1993; Rotonda et al., 1993) arrives at a similar conclusion in their analysis and comparison of the structures of human and yeast FKBP12–FK506 complexes and of the human FKBP12 complex with L-685,818, an 18-hydroxy,21-ethyl analog of FK506.

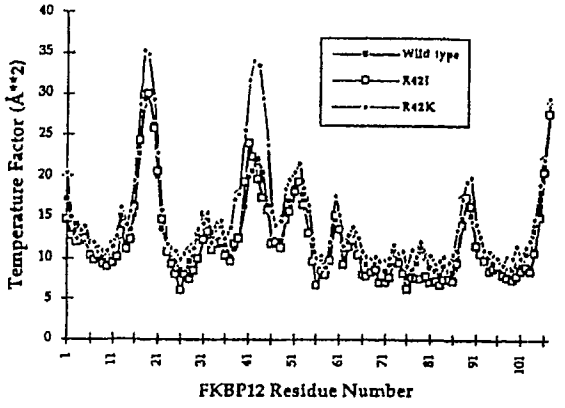


Fig. 4. Average X-ray temperature factors for the main-chain heavy atoms of FKBP12 for the native and the two mutant complex structures. Residues showing high temperature factors correspond to loop regions in the structure of FKBP12. These have been shown previously to be regions of local flexibility within the protein that were identified through a crystal structural analysis of 19 FKBP12–ligand complexes, each in a different crystal-packing arrangement (Wilson et al., 1995).

Yang et al. (1993), however, have suggested that substitutions at R42 exert their influence on CN inhibition indirectly, through a generalized conformational misorientation of nearby elements of the FKBP12–FK506 effector surface. This model is inferred from the curious pattern of CN-inhibitory activity evidenced in the FKBP12/13 chimeras studied by these workers. Substitution in FKBP12 of the corresponding 40s loop sequence from FKBP13 (i.e., replacing the sequence RDRNK with LPQNQ) leads to only a modest loss of CN-inhibitory potency (by ~2-fold to 19 nM). In turn, the single site R42Q mutant complex is severely compromised (by ~100-fold, to 850 nM), even though the R42Q substitution is incorporated in the FKBP12/13 chimera. From these observation, Yang et al. (1993) concluded that the effects of an R42 substitution would have to be strongly contextual. In other words, R42 and Q42 would each be appropriate to the 40s loop of FKBP12 and FKBP13 respectively, with only a modest loss of activity for the latter in the chimeric FKBP12/13 complex. An incompatible substitution, such as that of R42Q into FKBP12, would then lead to a significant and generalized disruption of the effector surface, an event that would be reflected in the much lower CN-inhibitory activity seen in the single-mutant complexes. Clardy (1995) has noted that the 40s loop in the structure of the native FKBP13–FK506 complex is displaced by about 2 Å RMS relative to the FKBP12–FK506 complex when these are overlapped, a point in support of the Yang et al. (1993) thesis.

The structural data presented here for the R42K and R42I mutant complexes show no such significant rearrangement of the 40s loop, the 80s loop, or any other part of the FKBP12 protein (Fig. 3). Nor do we observe a change in the conformation of FK506, even though we have demonstrated elsewhere (Itoh et al., 1995) that just such a conformational transformation is present in the FKBP12 R42K–H87V double mutant complex (Kinemage 2). All of our mutant complex structures (including the Itoh et al. [1995] double-mutant complex) have been solved in the same native FKBP12–FK506 complex crystal form described by van Duyne et al. (1991a), a crystal form that includes a significant number of ligand–ligand interactions (van Duyne et al., 1993; Wilson et al., 1995) that might otherwise have compromised the comparative interpretation we've presented for these structures. Even in the R42K mutant complex, where a significant increase in the temperature factors of the 40s loop of the mutant complex structure is observed (see Fig. 4), the weakened electron density in this region (Fig. 3C) is still consistent with the conformation of the 40s loop found in the native complex structure. In the less conservative R42I mutant complex structure, however, the fortuitous and unexpected ordering of two water molecules (Fig. 3A,B) acting as surrogates of the missing guanidino nitrogens of R42, leads to a more highly conserved structure, even though the loss of CN-inhibitory activity is also greater (~180-fold at 970 nM for the R42I mutant complex versus ~110-fold at 590 nM for the R42K mutant complex; Table 1).

Elsewhere, we describe the structures of FKBP12 (Itoh et al., 1995) and FKBP13 (Griffith JP, Wilson KP, Futer O, Living-

ston DJ, Navia MA. Structure of a mutant FKBP13-FK506 complex that is a high affinity inhibitor of calcineurin [manuscript in preparation]) mutant complexes that are inconsistent with the hypothesis that significant structural rearrangements in the 80s loop region of FKBP12 are responsible (Yang et al., 1993; Clardy, 1993, 1995) for the loss of CN-inhibitory activity seen in the corresponding FKBP12 mutant complexes. Nor do the results presented here support a similar hypothesis for the 40s loop region. Given our structural observations, one might speculate that R42 participates in some direct interaction with CN that cannot be effectively mimicked by the guanidino nitrogen surrogate water molecules that stabilize the 40s loop in the R42I mutant complex structure. With the R42K mutant complex, one might further consider an intermediate level of interaction with CN, through the single ϵ-amino nitrogen of the lysine substituent, and in spite of the greater disorder in the 40s loop. Teague (1995) has recently restated the importance of a conformationally well-defined recognition surface in promoting a strong interaction between exposed hydrophobic elements, such as are found in the FKBP12–FK506 effector surface and are presumed to exist on the complementary surface of CN. Our results show that relatively minor, localized perturbations of the CN complementarity of the FKBP12–FK506 effector surface can lead to quite significant effects on the CN-inhibitory potential of the resulting mutant complexes.

Methods

Mutant FKBP12 protein was prepared as reported (Aldape et al., 1992; Park et al., 1992; Wilson et al., 1995). Native and mutant FKBP12–FK506 complexes were prepared and crystals were grown essentially as described (van Duyne et al., 1991a; Wilson et al., 1995). Native crystals were used to seed the mutant complex crystallization experiments, and all the species reported here crystallized isomorphously in the native crystal form, as shown in Table 2. Diffraction data were collected on an X1000 multiwire area detector (Siemens Analytical Instruments, Madison, Wisconsin) or on an R-Axis II image plate detector (Rigaku/MSC, Woodlands, Texas), as indicated in Table 2. Data collection and processing used software provided by the manufacturers. All data were collected at room temperature. The reported structure of native FKBP12 in complex with FK506 (van Duyne et al., 1991a; Brookhaven Protein Data Bank [Bernstein et al., 1977] entry 1FKF) was used directly as an initial model for the crystallographic refinement of the mutant and native protein complexes. Refinement was by simulated annealing using the X-PLOR program package (Brünger, 1992). Mutated amino acids were initially refined as alanine, and the actual mutant side chains were introduced as refinement progressed. Water molecules were positioned in the model with the aid of a peak search program (SERC Daresbury Laboratory, 1979). The program QUANTA (Molecular Simulations; Burlington, Massachusetts) was used to examine electron density maps and protein models and for the superposition of structures and the calculation of the RMS differences reported (Table 2). Coordinates for the structures are being deposited in the Protein Data Bank.

Acknowledgments

We thank our colleagues Robert A. Aldape, Steven Chambers, Paul Caron, Maureen DeCenzo, John Fulghum, Olga Futer, James Griffith, David Livingston, and Mark Murcko for their continued assistance and support. In particular, Maureen DeCenzo and David Livingston provided the native and mutant proteins used in these studies. Paul Caron prepared the Kinemage files that accompany this paper. We also thank Joshua Boger, Patrick Connelly, Christopher Lepre, and Vicki Sato for their comments on this manuscript.

References

Alber T, Dao-pin S, Wilson K, Wozniak JA, Cook SP, Matthews BW. 1987. Contributions of hydrogen bonds of Thr 157 to the thermodynamic stability of phage T4 lysozyme. *Nature (Lond) 330*:41–46.

Aldape RA, Futer O, DeCenzo MT, Jarrett BP, Murcko MA, Livingston DJ. 1992. Charged surface residues of FKBP12 participate in formation of the FKBP12–FK506 calcineurin complex. *J Biol Chem 267*:16029–16032.

Armistead DA, Badia MC, Deininger DD, Duffy JP, Saunders, JO, Tung RD, Thomson JA, DeCenzo MT, Futer O, Livingston DJ, Murcko MA, Yamashita MM, Navia MA. 1995. Design, synthesis and structure of nonmacrocyclic inhibitors of FKBP12, the major binding protein for the immunosuppressant FK506. *Acta Crystallogr D51*:522–528.

Becker JW, Rotonda J, McKeever BM, Chan HK, Marcy AI, Wiederrecht G, Hermes JD, Springer JP. 1993. FK-506-binding protein: Three dimensional structure of the complex with the antagonist L-685,818. *J Biol Chem 268*:11335–11339.

Bernstein FC, Koetzel TF, Williams GJB, Meyer EF Jr, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M. 1977. The Protein Data Bank: A computer-based archival file for macromolecular structures. *J Mol Biol 112*:535–542.

Bierer BE, Mattila PS, Standaert RF, Herzenberg LA, Burakoff SJ, Crabtree G, Schreiber SL. 1990a. Two distinct signal transmission pathways in T-lymphocytes are inhibited by complexes formed between an immunophilin and either FK506 or rapamycin. *Proc Natl Acad Sci USA 87*: 9231–9235.

Bierer BE, Somers PK, Wandless TJ, Burakoff SJ, Schreiber SL. 1990b. Probing immunosuppressant action with a nonnatural immunophilin ligand. *Science 250*:556–559.

Brünger AT. 1992. *X-PLOR version 3.1: A system for X-ray crystallography and NMR*. New Haven, Connecticut: Yale University.

Clardy J. 1993. Structural studies of complexed FK-506 binding protein. *Ann NY Acad Sci 685*:37–46.

Clardy J. 1995. The chemistry of signal transduction. *Proc Natl Acad Sci USA 92*:56–61.

Connelly PR, Aldape RA, Bruzzese FJ, Chambers SP, Fitzgibbon MJ, Fleming MA, Itoh S, Livingston DJ, Navia MA, Thomson JA, Wilson KP. 1994. Enthalpy of hydrogen bond formation in a protein–ligand binding reaction. *Proc Natl Acad Sci USA 91*:1964–1968.

Dumont FJ, Melino MR, Staruch MJ, Koprak SL, Fischer PA, Sigal NH. 1990a. The immunosuppressive macrolides FK-506 and rapamycin act as reciprocal antagonists in murine T cells. *J Immunol 144*:1418–1424.

Dumont FJ, Staruch MJ, Koprak SL, Melino MR, Sigal NH. 1990b. Distinct mechanisms of suppression of murine T cell activation by the related macrolides FK-506 and rapamycin. *J Immunol 144*:251–258.

Friedman J, Weissman I. 1991. Two cytoplasmic candidates for immunophilin action are revealed by affinity for a new cyclophilin; one in the presence and one in the absence of CsA. *Cell 66*:799–806.

Futer O, DeCenzo MT, Aldape RA, Livingston DJ. 1995. FK506 binding protein mutational analysis: Defining the surface residue contributions to stability of the calcineurin co-complex. *J Biol Chem 270*:18935–18940.

Goulet MT, Rupprecht KM, Sinclair PJ, Wyvratt MJ, Parsons WH. 1994. The medicinal chemistry of FK-506. *Perspect Drug Disc Design 2*:145–162.

Itoh S, DeCenzo MT, Livingston DJ, Pearlman DA, Navia MA. 1995. Conformation of FK506 in X-ray structures of its complexes with human recombinant FKBP12 mutants. *Bioorg Med Chem Lett 5*:1983–1988.

Kino T, Hatanaka H, Hashimoto M, Nishiyama M, Goto T, Okuhara M, Kohsaka M, Aoki H, Imanaka H. 1987. FK-506, a novel immunosuppressant isolated from a *Streptomyces*. I. fermentation, isolation and physico-chemical and biological characteristics. *J Antibiot (Tokyo) 40*: 1249–1255.

Klee CB, Cohen P. 1988. The calmodulin-regulated protein phosphatase. *Mol Aspects Cell Regul 5*:225–248.

Liu J, Farmer JD Jr, Lane WS, Friedman J, Weissman I, Schreiber SL. 1991. Calcineurin is a common target of cyclophilin–cyclosporin A and FKBP–FK506 complexes. *Cell 66*:807–815.

Matthews BW. 1993. Structural and genetic analysis of protein stability. *Annu Rev Biochem 62*:139–160.

Park ST, Aldape RA, Futer O, DeCenzo MT, Livingston DJ. 1992. PPIase

catalysis by human FK506-binding protein proceeds through a conformational twist mechanism. *J Biol Chem 267*:3316–3324.
Rosen MK, Schreiber SL. 1992. Natural products as probes of cellular function: Studies on immunophilins. *Angew Chem Intl Ed Engl 31*:384–400.
Rotonda J, Burbaum JJ, Chan KH, Marcy AI, Becker JW. 1993. Improved calcineurin inhibition by yeast FKBP12-drug complexes: Crystallographic and functional analysis. *J Biol Chem 268*:7607–7609.
Schreiber SL. 1991. Chemistry and biology of the immunophilins and their immunosuppressive ligands. *Science 251*:283–287.
Schreiber SL, Albers MW, Brown EJ. 1993. The cell cycle, signal transduction, and immunophilin–ligand complexes. *Acc Chem Res 26*:412–420.
Schultz LW, Martin PK, Liang J, Schreiber SL, Clardy J. 1994. Atomic structure of the immunophilin FKBP13-FK506 complex: Insights into the composite binding surface for calcineurin. *J Am Chem Soc 116*:3129–3130.
Sehgal SN, Baker H, Vezina C. 1975. Rapamycin (AY-22,989), a new antifungal antibiotic. I. Taxonomy of the producing streptomycete and isolation of the active principal. *J Antibiot (Tokyo) 28*:727–732.
SERC Daresbury Laboratory. 1979. *CCP4: Collaborative computing project no. 4. A suite of programs for protein crystallography.* Warrington WA4 4AD, UK: Daresbury Laboratory.
Somers PK, Wandless TJ, Schreiber SL. 1991. Synthesis and analysis of 506BD, a high-affinity ligand for the immunophilin FKBP. *J Am Chem Soc 113*:8045–8056.
Teague S. 1995. Lessons from molecular matchmakers. *Struct Biol 2*: 360–361.
van Duyne GD, Standaert RF, Karplus PA, Schreiber SL, Clardy J. 1991a. Atomic structure of FKBP–FK-506, an immunophilin–immunosuppressant complex. *Science 252*:839–842.
van Duyne GD, Standaert RF, Karplus PA, Schreiber SL, Clardy J. 1993. Atomic structures of human immunophilin FKBP12 complexes with FK506 and rapamycin. *J Mol Biol 229*:105–124.
van Duyne GD, Standaert RF, Schreiber SL, Clardy J. 1991b. Atomic structure of the rapamycin human immunophilin FKBP-12 complex. *J Am Chem Soc 113*:7433–7434.
Wilson KP, Yamashita MM, Sintchak MD, Rotstein SH, Murcko MA, Boger J, Thomson JA, Fitzgibbon MJ, Black JR, Navia MA. 1995. Comparative X-ray structures of the major binding protein for the immunosuppressant FK506 (tacrolimus) in unliganded form and in complex with FK506 and rapamycin. *Acta Crystallogr D 51*:511–521.
Yang D, Rosen MK, Schreiber SL. 1993. A composite FKBP12–FK506 surface that contacts calcineurin. *J Am Chem Soc 115*:819–820.

Note added in proof

A structure determination of the native FKBP12-FK506 complex bound to calcineurin has now been reported (Griffith JP, Kim JL, Kim EE, Sintchak MD, Thomson JA, Fitzgibbon MJ, Fleming MA, Caron PR, Hsiao K, Navia MA. 1995. X-ray structure of calcineurin inhibited by the immunophilin-immunosuppressant FKBP12-FK506 complex. *Cell 82*:507–522). A preliminary fit of mutant FKBP12 complex structures to the native FKBP12 in the calcineurin complex is entirely consistent with the results presented in this paper.

THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 267, No. 4, Issue of February 5, pp. 2393-2399, 1992
Printed in U.S.A.

Hopanoid B

Tolerance of T4 Lysozyme to Proline Substitutions within the Long Interdomain α-Helix Illustrates the Adaptability of Proteins to Potentially Destabilizing Lesions*

(Received for publication, September 13, 1991)

Uwe H. Sauer‡, Sun Dao-pin§, and Brian W. Matthews¶

From the Institute of Molecular Biology, Howard Hughes Medical Institute and Department of Physics, University of Oregon, Eugene, Oregon 97403

To investigate the ability of a protein to accommodate potentially destabilizing amino acid substitutions, and also to investigate the steric requirements for catalysis, proline was substituted at different sites within the long α-helix that connects the amino-terminal and carboxyl-terminal domains of T4 lysozyme. Of the four substitutions attempted, three yielded folded, functional proteins. The catalytic activities of these three mutant proteins (Q69P, D72P, and A74P) were 60–90% that of wild-type. Their melting temperatures were 7–12 °C less than that of wild-type at pH 6.5. Mutant D72P formed crystals isomorphous with wild-type allowing the structure to be determined at high resolution. In the crystal structure of wild-type lysozyme the interdomain α-helix has an overall bend angle of 8.5°. In the mutant structure the introduction of the proline causes this bend angle to increase to 14° and also causes a corresponding rotation of 5.5° of carboxyl-terminal domain relative to the amino-terminal one. Except for the immediate location of the proline substitution there is very little change in the geometry of the interdomain α-helix. The results support the view that protein structures are adaptable and can compensate for potentially destabilizing amino acid substitutions. The results also suggest that the precise shape of the active site cleft of T4 lysozyme is not critical for catalysis.

Phage T4 lysozyme is a small monomeric protein with its structure divided into two distinct domains (Fig. 1). The active site is located at the junction of the two domains, and it might be expected that the alignment of one domain relative to the other would be critical for catalytic activity (*cf.* Storm and Koshland, 1970, but see also Jenks, 1969; Knowles, 1991). On the other hand, the crystal structure of a fully active mutant of T4 lysozyme has recently been described in which there is substantial variability in the "hinge-bending angle" between one domain and the other (Faber and Matthews, 1990).

To investigate the ability of T4 lysozyme to compensate for disruptive changes in its structure and to determine the need to conserve the alignment of the active site cleft, a series of proline substitutions was made in the long interdomain α-helix. Studies of proline substitutions and proline replacements in proteins and peptides include Matthews *et al.* (1987), Alber *et al.* (1988), O'Neil and DeGrado (1990), Strehlow *et al.* (1991), and Consler *et al.* (1991), among others.

The amino acids chosen for substitution with proline, Gln-69, Val-71, Asp-72 and Ala-74, are located in the middle of the long α-helix (residues 60–80) that connects the two domains of T4 lysozyme (Fig. 1). It was expected that the substitution of a proline at any of these sites would tend to significantly distort the α-helix and therefore change the alignment of the "upper" and "lower" domain. By making a series of replacements it was anticipated that the active-site cleft would be distorted in different ways. Also by including different substitutions it was possible to include sites that were both buried and solvent-exposed.

Gln-69 is largely exposed to solvent (Table I, Fig. 1) and its side chain does not obviously participate in stabilizing interactions with other parts of the protein. Val-71 is almost but not entirely buried. Its side chain makes many contacts with the non-polar residues Ile-3, Phe-4, Leu-7, Phe-67, Ala-74, Ile-100, and Phe-104 that contribute to the hydrophobic core both within the carboxyl-terminal domain of T4 lysozyme and connecting one domain with the other. Asp-72 is very solvent exposed and its side chain is located on the "backside" of the interdomain α-helix relative to the rest of the lysozyme molecule (Fig. 1). Ala-74 is largely, but not entirely buried, and makes non-polar contacts with residues Val-71, Ile-100, Val-103, and Phe-104 that contribute to the hydrophobic core in the carboxyl-terminal lobe of the molecule. In wild-type lysozyme Asp-70 makes an unusually strong salt bridge with His-31 (Anderson *et al.*, 1990). Asp-70 also accepts a hydrogen bond from the backbone amide of Ala-74. Since the replacement of Ala-74 with proline eliminates any possibility of hydrogen bonding to the amide it was likely that this replacement would perturb the Asp-70. . .His-31 salt bridge.

EXPERIMENTAL PROCEDURES

Mutagenesis—The proline mutations were introduced by site-directed mutagenesis according to the uracil template method developed by Kunkel (Kunkel, 1985; Kunkel *et al.*, 1987). The cysteine-free "pseudo wild-type" lysozyme (WT*),[1] in which the 2 cysteine residues present in wild-type had been replaced in order to facilitate thermodynamic measurements (Wetzel *et al.*, 1988; Matsumura and Matthews, 1989; Pjura *et al.*, 1990) was used as the reference protein. The lysozyme *e-gene contained within a 630 base-pair BamHI-HindIII* fragment had been previously cloned into phage M13 mp18 yielding the derivative M13 mp18 T4e C54T C97A. It was then transformed into *Escherichia coli* strain CJ236 (dut, ung-, thi1, relA/pCJ105(CM'))

* This work was supported in part by National Institutes of Health Grant GM21967 and the Lucille P. Markey Charitable Trust. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ Present address: European Molecular Biology Laboratory, Postfach 10.2209, Meyerhofstrasse 1, W-6900 Heidelberg, Germany.

§ Present address: Laboratory of Molecular Biology, NIDDKD, National Institutes of Health, Bldg. 2, Rm. 316, Bethesda, MD 20892.

¶ To whom correspondence should be addressed.

[1] The abbreviation used is: WT*, pseudo wild-type.

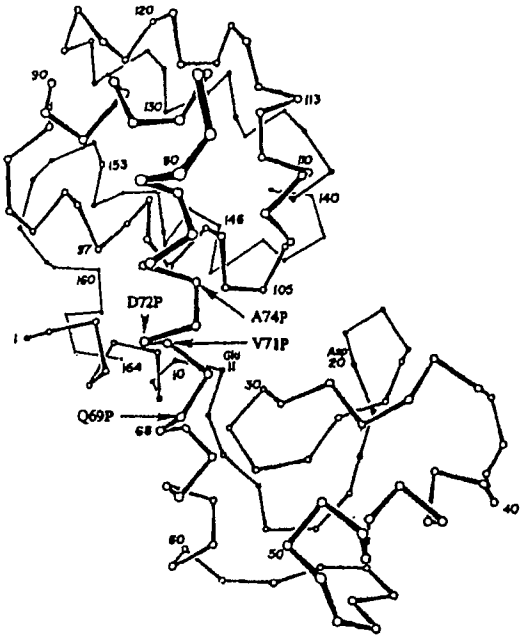


FIG. 1. **Backbone of T4 lysozyme showing the locations of the proline substitutions discussed in the text.** Mutations Q69P, D72P, and A74P give active, folded protein.

which was used to prepare uracil-containing single-stranded template DNA. After annealing of the mutagenic primer to WT* template DNA, circularization was accomplished by using the Klenow fragment of *E. coli* DNA polymerase I and T4 DNA ligase. The double-stranded DNA was subsequently used to transform competent *E. coli* JM101 cells. Sequencing of the whole lysozyme gene was carried out to confirm that no changes had occurred other than the ones introduced.

Protein Purification—Following subcloning into the plasmid expression vector pHN1403 (Muchmore *et al.*, 1989; Poteete *et al.*, 1991) 100 ml of cells grown overnight in 100 ml of LBH broth (10 g of tryptone, 5 g of NaCl, 1 ml of 1 M NaOH/liter) was added to 3 liters of LB broth (12 g of tryptone, 5 g of yeast extract, 10 g of NaCl, 1 g of glucose/liter) and grown at 37 °C under constant agitation (700 rpm) and air flow (12 liters/min) in a 5-liter fermenter until the optical density at a wavelength of 595 nm reached a value of 1.2. The temperature was then decreased to 26 °C, agitation and aeration were reduced to 200 rpm and 7 liters/min, respectively. Isopropyl-β-thiogalactoside (800 mg) was added to the growth media in order to induce lysozyme expression which was allowed to proceed for 100 min under continued stirring and aeration. The cells were then harvested into a 5-liter Erlenmeyer flask where they lysed. A few grains of DNase I were added to the now thick and viscous lysate which was left stirring at 4 °C for 2.5 h. The lysed cell suspension regained almost the viscosity of LB broth, was then taken out to room temperature, and placed on a magnetic stirrer for 30 min to allow for complete cell lysis. This step almost doubled the final yield of proline-containing mutant lysozymes. The lysate became more viscous again and was placed back at 4 °C, stirring for another 1.5 h until it regained the fluidity of LB broth. All the subsequent purification steps were carried out at 4 °C. After centrifugation at 10,000 rpm (17,700 × *g*) for 2 h, only the supernatant contained mutant lysozyme. It was dialyzed against 20 mM phosphate buffer at pH 6.5 until the conductivity reached a value of 3.8 mS/cm (1 Siemens = 1 Ω^{-1}). The dialyzed supernatant was loaded onto a 2.5 × 10-cm CM-Sepharose ion-exchange column which had previously been equilibrated with 400 ml of 50 mM Tris buffer at pH 7.3. Mutant lysozyme was eluted using an 800-ml salt gradient in the range from 50 to 300 mM NaCl in 50 mM Tris buffer, pH 7.3. Elution was monitored at a wavelength of λ = 280 nm. Peak fractions (absorbance at a wavelength of 280 nm above 0.4 units) were pooled, dialyzed against 50 mM phosphate buffer at pH 5.8 for 12 h, and concentrated on a 1 × 5-cm SP-Sephadex column previously equilibrated with the same buffer. The protein was eluted with SP buffer (100 mM $NaPO_4$, pH 6.5, 550 mM NaCl, 0.02% NaN_3) and stored in this buffer at 4 °C. Based on sodium dodecyl sulfate and high performance liquid chromatography analysis the purity of the mutant T4 lysozyme was estimated to be over 95%. Final amounts of 60–120 mg of mutant T4 lysozyme proteins were obtained. Wild-type T4 lysozyme typically yields about 150 mg/3-liter preparation.

The activity of the mutant lysozyme was measured at room temperature using the turbidity assay described by Tsugita (Tsugita *et al.*, 1968). Since absolute rates were not reproducible, activities were normalized to a wild-type control.

Measurement of Thermal Stability—Thermal denaturations at pH 2.0 and 6.5 were monitored by circular dichroism (CD) at a wavelength of λ = 229 nm on a Jasco J-500C spectropolarimeter as a function of temperature (Elwell and Schellman, 1975; Dao-pin *et al.*, 1990). The temperature was varied from 0 to 75 °C at a rate of 1 °C/min with a Hewlett-Packard 89100A temperature controller interfaced to a Hewlett-Packard 87 XM computer. The protein concentration was adjusted to 0.02 mg/ml by measuring the absorbance at λ = 280 nm using a double beam Varian 2290 spectrophotometer. A probe immersed in the sample solution just above the UV beam recorded the temperature. The solution was continuously stirred with a magnetic stirbar. Buffer solutions (150 mM KCl, 10 mM HCl at pH 2.0 and 150 mM KCl, 10 mM potassium phosphate at pH 6.5) were prepared from doubly deionized, degassed H_2O and were filtered before use through a 22-μm Millipore filter unit.

Thermal denaturations were repeated at least three times. Measurements of the mutants were flanked by WT* thermal denaturations under exactly the same conditions in order to minimize errors (Dao-pin *et al.*, 1990). The data were analyzed using standard van't Hoff techniques (Becktel and Schellman, 1987; Dao-pin *et al.*, 1990).

Crystallographic Methods—Crystal growth was attempted using both hanging-drop[2] as well as batch methods (Weaver and Matthews, 1987; Alber and Matthews, 1987) under conditions similar to those used for wild-type lysozyme. Both methods yielded crystals of D72P isomorphous with WT in the space group $P3_221$.

Refinement was carried out using the TNT package of refinement programs (Tronrud *et al.*, 1987). The positional coordinates and the temperature factors were refined simultaneously using the "conjugate directions" option in TNT which improves convergence.[3]

The starting model for refinement was the refined structure of the cysteine-free wild-type (WT*) (Pjura *et al.*, 1990; Bell *et al.*, 1991)[2] with residue 72 truncated to Ala. The general approach was to begin with low resolution (8–4 Å) rigid body refinement with the molecule considered as a single unit. This was followed by further rigid body refinement but with the mutant molecule divided into two parts (residues 1–80 and 81–162). Finally, the molecule was divided into three blocks (residues 1–59, 60–80, and 81–162). After examining the model on the graphics terminal (Jones, 1978), proline was built at position 72 and water molecules from the WT* structure were included in the model. Several cycles of positional refinement using moderately weighted geometric constraints were performed using data between 20 and 1.9 Å. Again, the model was inspected on the graphics terminal, some water molecules were added, others repositioned, and some side chains adjusted to better fit the electron density. Only those water molecules which had a final refined temperature factor of less than 80 $Å^2$, and, in addition, formed hydrogen bonds to protein or other bound water molecules and had no steric clashes were retained. Thereafter, several cycles of simultaneous positional and temperature factor refinement were alternated with model building until the crystallographic residual converged. The number of solvent molecules in the refined model is roughly the same as for WT* lysozyme, *i.e.* about 145. The refined coordinates have been deposited in the Brookhaven Data Bank.

RESULTS

Expression of Mutant Proteins

Gln-69 → Pro—Mutant Q69P could be expressed and purified in a straightforward manner, yielding up to 100 mg of protein from 4 liters of culture medium. The activity of the protein is close to that of wild-type (Table I), but it is less stable, with melting temperature 12.9 °C less than wild-type at pH 2.2 and 7.6 °C lower at pH 6.5. This behavior corresponds to that of a typical "temperature sensitive" mutant of

[2] A. E. Eriksson, W. A. Baase, and B. W. Matthews, manuscript in preparation.

[3] D. E. Tronrud, submitted for publication.

TABLE I
Sites of proline replacement

Amino acid	Fraction of side chain accessible to solvent	Mutant	Relative activity
			%
Gln-69	0.75	Q69P	88
Val-71	0.06	V71P	*
Asp-72	0.76	D72P	57
Ala-74	0.04	A74P	66

* Purified protein was not obtained for V71P.

TABLE II
Thermal stability of proline-containing mutants

T_m is the melting temperature and ΔT_m the difference between the melting temperature of the mutant and that of the pseudo wild-type lysozyme (see text). $\Delta\Delta G$, the difference between the free energy of unfolding of the mutant and pseudo wild-type lysozyme, was estimated from the relationship $\Delta\Delta G = \Delta S \cdot \Delta T_m$ (Becktel and Schellman, 1987) where ΔS is the entropy of unfolding of the wild-type protein (257 and 378 cal/degree mol at pH 2.0 and 6.5). This relationship may not be reliable when ΔT_m is large. The quoted values of $\Delta\Delta G$ are subject to error, estimated as ±0.5 kcal/mol for Q69P and D72P and ±1 kcal/mol for A74P. The estimated error in T_m is ±0.3 °C for Q69P and D72P and ±0.5 °C for A74P.

Protein	pH	T_m of mutant	T_m of WT*	ΔT_m	$\Delta\Delta G$
			°C		kcal/mol
Q69P	2.0	25.6	38.5	−12.9	−3.3
	6.5	55.6	63.2	−7.6	−2.9
D72P	2.0	28.3	38.5	−10.2	−2.6
	6.5	56.3	63.4	−7.1	−2.7
A74P	3.5	39.3	57.5	−18.2	−5.7
	4.0	46.9	62.2	−15.3	−5.0
	5.5	53.3	65.7	−12.4	−4.5
	6.5	50.8	62.9	−12.1	−4.6

T4 lysozyme selected by the random screen of Streisinger *et al.* (1961) (*cf.* Grütter *et al.*, 1979, 1987; Hawkes *et al.*, 1984). When stored at 4 °C at about 50 mg/ml, Q69P tended to form white opalescent aggregates which dissolved on warming to room temperature. This process was reversible and seemed to have no effect on stability or activity. Small crystals of the protein were obtained, apparently non-isomorphous with wild-type.

Val-71 → Pro—When V71P DNA was transformed into *E. coli* and the ability of a bacterial extract tested to form a halo on isopropyl-1-thio-β-D-galactopyranoside lysis indicator plates, no halo could be seen at 37 °C. After 24 h at 4 °C, a small halo was visible but we cannot rule out the possibility that this might be due to a small amount of WT* present as an impurity. Attempts to purify V71P by the method described above, or by using a French press to break open the bacterial cell walls, or by an alternative method described by Dao-pin *et al.* (1991b) were all unsuccessful. We presume that V71P is very unstable and/or is rapidly degraded by proteolysis.

Asp-72 → Pro—As with Q69P this protein could be readily purified by the standard procedure, yielding up to 130 mg of protein from a 4-liter culture. Its activity is about 60% that of wild-type (Table I) and it is less stable (Table II), again, roughly comparable to a typical temperature-sensitive mutant. Crystals isomorphous with wild-type could be grown at 4 °C using both batch and hanging-drop techniques. The latter method gave the largest crystals, 0.5 × 0.65 × 0.3 mm, from 2.0 M phosphate at pH 7.1. Some apparently non-isomorphous crystals were also obtained at 15 °C but have not been examined.

Ala-74 → Pro—As for Q69P and D72P, A74P behaved normally and yielded ~140 mg of protein/preparation from a 4-liter culture. The apparent activity is about two-thirds that of wild-type (Table 2). This mutant is less stable than both D72P and Q69P (Table II). Because of the low stability the protein tends to be partially unfolded at low pH, so the stability measurements under these conditions are less reliable than at higher pH. Some crystals of this protein were obtained from 50 mM phosphate, pH 7.9, 16% PEG 8000 and are non-isomorphous with wild-type.

Structure of D72P

The crystals of D72P appeared to be somewhat more sensitive to radiation than wild-type lysozyme and did not diffract as well. For this reason the exposure time per frame on a Xuong-Hamlin area detector system operating with graphite-monochromated CuK_α radiation from a Rigaku generator (40 kV, 100 mA) was increased to 60 s, compared to the usual 30 s for WT. Under these conditions the data set to 1.9 Å was 89% complete (Table III).

The map showing the difference in density between D72P and WT* lysozyme is shown in Fig. 2*a*. The positive density feature confirms the addition of the pyrrolidine ring. Also there is negative density at the site previously occupied by the carboxylate of Asp-72. Positive and negative densities also indicate the movement of the carbonyl oxygen of Asn-68 away from the helix axis (Fig. 2*b*). A negative feature next to the carbonyl group of Asp-70 indicates a movement of the oxygen toward the helix axis. The side chain of Asp-70 also moves slightly, as does His-31 (not shown) maintaining the strong (Anderson *et al.*, 1990) His-31...Asp-70 salt bridge. The movement of His-31 together with slight adjustments occurring throughout the lower lobe give the impression that the lower part of the long α-helix and the lower domain move as an essentially connected unit.

There is, however, a movement of the upper domain relative to the lower one. This shift is shown in Figs. 3 and 4. If backbone atoms in the amino-terminal domain of the mutant structure (residues 13–59) are superimposed on the corresponding atoms in WT* lysozyme (Fig. 4*a*), they have root-mean-square discrepancy of 0.16 Å, which is essentially experimental error. Similarly, the respective backbone structures within the carboxyl-terminal domains are also well conserved (root-mean-square discrepancy of 0.18 Å for residues 81–162, Fig. 4*b*). This shows that the structures within

TABLE III
Data collection and refinement statistics
Data for WT* taken from Eriksson *et al.*[2]

Protein	WT* (C54T/C97A)	D72P
Data collection statistics		
Mode of data acquisition	Film	Area detector (Xuong-Hamlin)
Cell dimensions		
a, b (Å)	60.9	60.8
c (Å)	96.8	98.6
Resolution (Å)	1.75	1.9
Unique reflections	14,562	15,147
Completeness of data (%)	65.0	89.4
R_{merge} (on intensities) (%)	4.6	5.5
Refinement statistics		
Resolution limits (Å)	6.0–1.75	20.0–1.9
RMS* deviation from ideal values		
Bond length (Å)	0.015	0.013
Bond angle (°)	2.1	2.0
Crystallographic residual (%)	14.8	16.7

* Root-mean-square.

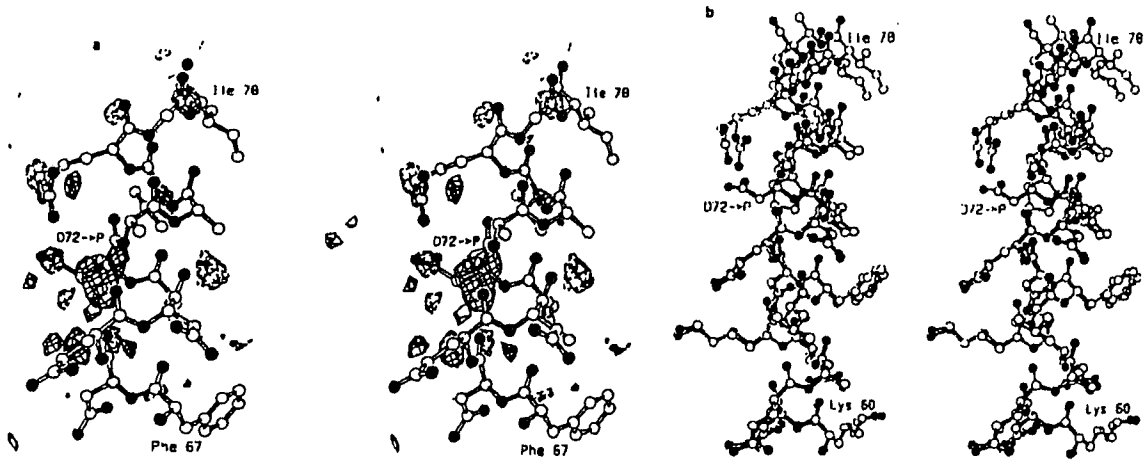


FIG. 2. *a*, map showing the difference in density between D72P lysozyme and wild-type. Amplitudes ($F_{Mut} - F_{WT}$) where F_{Mut} and F_{WT} are the structure amplitudes observed for the mutant and pseudo wild-type crystals. Phases calculated from the refined pseudo wild-type structure. Positive contours (*solid*) and negative contours (*broken*) drawn, respectively, at $+3\sigma$ and -3σ where σ is the root-mean-square value of the density throughout the unit cell. *b*, superposition of the A72P mutant structure (*solid bonds*) on wild-type (*open bonds*) in the vicinity of the mutation. The superposition of the coordinates shown were optimized by least squares prior to drawing the figure.

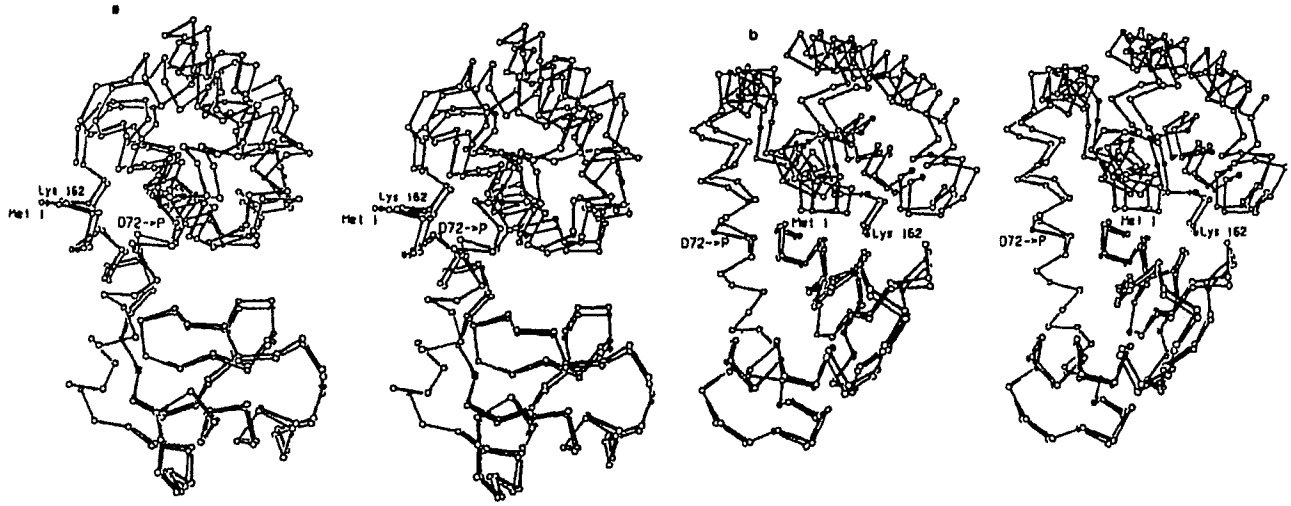


FIG. 3. *a*, backbone of the D72P mutant molecule (*dark bonds*) superimposed on WT* (*open bonds*). The figure is based on the optimal superposition of the amino-terminal region (residues 60–66) of the interdomain α-helix. *b*, superposition of D72P (*solid bonds*) on WT* (*open bonds*), as in Fig. 3*a* but rotated 90°.

the respective NH_2-terminal and COOH-terminal domains are conserved. It is the alignment of one domain relative to the other that changes in the mutant structure relative to WT*. As shown in Fig. 4, *a* and *b*, this movement corresponds to atom shifts up to about 1.5 Å. The pronounced maxima and minima in Fig. 4, *a* and *b*, are due to different distances of the corresponding atoms from the axis of rotation.

One way to analyze for movements of one part of the structure relative to another is to align the mutant structure with wild-type based on the superposition of relatively short segments of backbone. Two such alignments, based on 6-residue backbone segments on either side of the mutation site, are shown in Fig. 5, *a* and *b*. In Fig. 5*a*, in which the alignment is based on residues 60–66, the amino-terminal domain of the mutant structure agrees well with that of WT*. This shows that residues 60–66 and the NH_2-terminal domain are connected essentially as a rigid body. In contrast, when the superposition is based on residues 74–80 (Fig. 5*b*), neither the NH_2-terminal domains nor the COOH-terminal domains of the mutant and wild-type structures coincide. This suggests that the residues 74–80 are not connected rigidly to the COOH-terminal domain. The upper part of the 60–80 α-helix appears to be at least in part responsible for the observed flexibility. Small changes in this region can have large effects on the rest of the structure. The regions at the beginning and at the end of the long helix have been previously identified as "hinge-bending" regions in the M6I variant of T4 lysozyme (Faber and Matthews, 1990). The hinge-bending that was observed for the M6I mutant was assumed to be a low energy displacement because different molecules within the same crystal displayed different hinge-bending angles. In the case of M6I, the long interdomain helix appeared to remain rigid. In contrast, in the present case the long interdomain helix is bent relative to wild-type. Therefore the hinge-bending seen for D72P (Fig. 3) has a very different origin, and presumably very different energetics, relative to that seen for M6I.

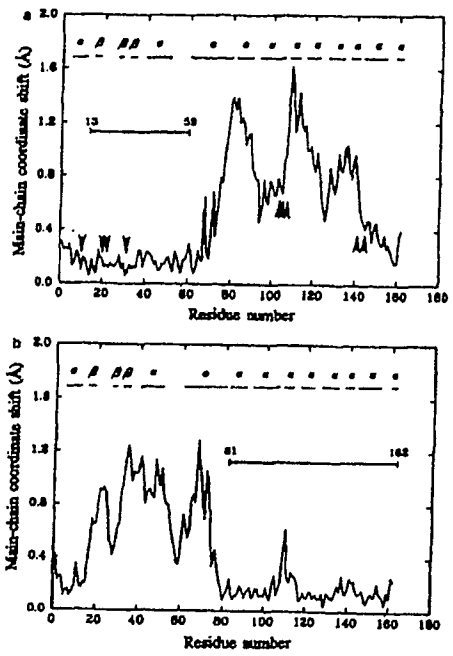


FIG. 4. **"Shift plots" showing the displacement between corresponding backbone atoms in the mutant D72P and wild-type lysozyme.** For each residue the value plotted is the root-mean-square discrepancy between the four backbone atoms in the mutant (N, C, CA, O) and the corresponding atoms in WT*. *a*, superposition of the structures of D72P and WT* based on residues 13–59 in the amino-terminal domain. The *arrowheads* indicate those residues that are presumed to contact an extended substrate (Glu-11, Asp-20, Thr-21, Leu-32, Phe-104, Gln-105, Thr-109, Thr-142 and Arg-145) (Anderson *et al.*, 1981). The contacts involving Thr-109 are to sugars in subsites A and B, whereas cleavage is between subsites D and E. *b*, superposition of D72P and WT* based on the carboxyl-terminal domain, residues 81–162.

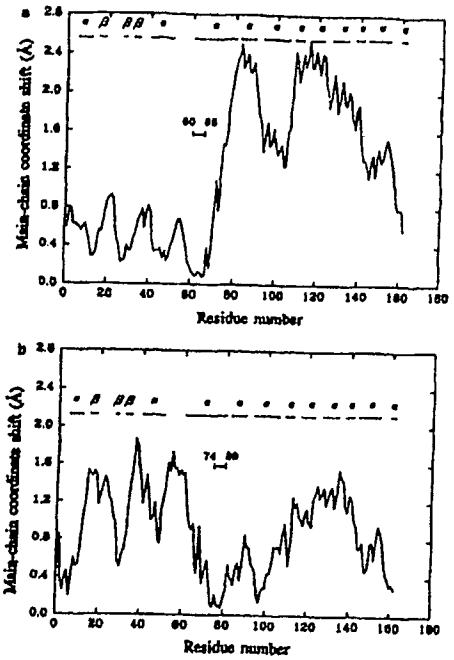


FIG. 5. *a*, superposition of D72P on WT* based on the alignment of the amino-terminal part of the interdomain α-helix, residues 60–66. *b*, superposition of D72P on WT* based on the alignment of the carboxyl-terminal part of the interdomain α-helix, residues 74–80. Note that *a* is similar to Fig. 4*a* indicating that residues 60–66 and the amino-terminal domain do not move very much relative to each other upon the proline replacement. *b*, however, is not very similar to Fig. 4*b*, showing that the carboxyl-terminal part of the interdomain α-helix does move somewhat relative to the carboxyl-terminal domain of the mutant protein.

Another way of detecting differences between two structures is by calculating a difference-distance matrix (Nishikawa *et al.*, 1972). All the intramolecular C_f^{α}-C_f^{α} distances are determined for one protein and then compared to the respective intramolecular distances of the second molecule. An example of such a plot is shown in Fig. 6. Since C_f^{α}-C_f^{α} distances in D72P were subtracted from the corresponding distances in WT*, positive contours (*solid lines*) reflect shorter distances in the mutant structure; negative contours (*dotted lines*) represent longer ones. As can be seen, the mutated residue 72 shows the largest shift with respect to residues 95, 122, and 157.

Fig. 2*b* illustrates the effect of the Asp-72 → Pro mutation on the long α-helix itself. In this figure residues 60–66 have been superimposed so as to make the structural change more obvious. The helix containing the proline has an overall bend of about 14°. A similar analysis of the structure of wild-type lysozyme, however, reveals that the same helix already has a bend of about 8.5°. Therefore the effect of the Asp-72 → Pro replacement is to increase the bending by about 5.5°.

Fig. 7 compares the hydrogen-bond distances within the interdomain helix in wild-type lysozyme and in the mutant structure. The first α-helical hydrogen bond is from the carbonyl oxygen of Thr-59 to the backbone amide of Ala-63. Not surprisingly, the introduction of the pyrrolidine ring at residue 72 results in a substantial increase in the distance between the nitrogen of residue 72 and the carbonyl oxygen of Asn-68. The nitrogen-oxygen distances for the successive residues are lengthened somewhat, but not excessively so (maximum 3.3 Å), suggesting that the hydrogen bonds within

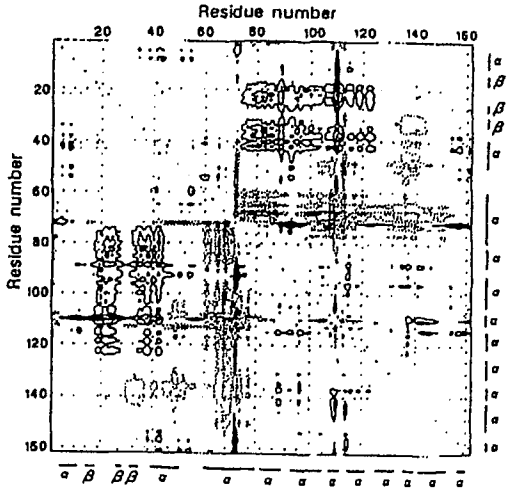


FIG. 6. **Difference distance matrix comparing distances between all pairs of α-carbon atoms in the mutant structure with the corresponding distances in WT*.** The quantity plotted is $\Delta_{ij} = r_{ij,WT^*} - r_{ij,D72P}$ where r_{ij,WT^*} is the distance between the *i*th and *j*th α-carbon atoms in the structure of wild-type lysozyme, and $r_{ij,D72P}$ is the distance between the *i*th and *j*th α-carbon atoms in the D72P mutant structure. *Solid contours* are drawn at 0.3 Å, 0.6 Å, 0.9 Å, ... and indicate pairs of α-carbon atoms that are closer together in the mutant structure than in wild-type. *Broken contours*, drawn at −0.3 Å, −0.6 Å, −0.9 Å, ... indicate pairs of α-carbon atoms that are further apart in the mutant than in wild-type. Featureless regions indicate domains within which the structure in the mutant is essentially identical with that in wild-type.

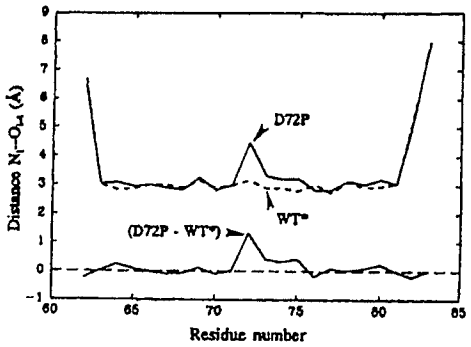


FIG. 7. **Hydrogen bond lengths within the interdomain helix.** The *upper plot* shows the distance between the nitrogen atom of residue i and the carbonyl oxygen of residue i-4. Within the α-helix (i = 63–81) this distance corresponds to a helical hydrogen bond. Values for the mutant, D72P, are indicated by the *solid line*; values for WT* are indicated by the *broken line*. The *lower plot* shows the difference in the distance between the mutant and wild-type structures.

TABLE IV

Ramachandran angles for residues within the long α-helix of T4 lysozyme

	D72P		WT*		Δφ	Δψ
	φ	ψ	φ	ψ		
Ile-58	−121.3	169.3	−126.2	167.2	4.8	2.1
Thr-59	−94.3	171.6	−90.9	166.4	−3.4	5.2
Lys-60	−62.9	−32.6	−56.8	−43.3	−6.1	10.7
Asp-61	−68.0	−35.9	−58.0	−45.7	−10.0	9.8
Glu-62	−69.1	−36.4	−62.3	−43.9	−6.8	7.5
Ala-63	−64.1	−41.9	−59.3	−46.1	−4.8	4.2
Glu-64	−70.4	−31.5	−69.5	−30.0	−1.0	−1.5
Lys-65	−61.1	−49.1	−69.1	−42.3	−8.1	−6.8
Leu-66	−61.3	−43.2	−59.4	−42.0	−2.0	−1.2
Phe-67	−55.5	−44.8	−61.6	−47.5	−6.1	−2.7
Asn-68	−62.4	−22.8	−56.2	−43.4	−6.2	20.6
Gln-69	−89.9	−38.9	−64.8	−40.1	−25.1	1.3
Asp-70	−66.6	−38.8	−67.8	−37.5	1.2	−1.4
Val-71	−63.8	−54.5	−68.1	−44.4	4.3	−10.1
Pro-72	−59.4	−34.8	−55.9	−48.8	−3.5	14.0
Ala-73	−64.1	−36.0	−60.7	−43.3	−3.4	7.3
Ala-74	−59.9	−50.2	−57.1	−56.2	−2.9	6.0
Val-75	−56.6	−48.0	−53.5	−49.7	−3.1	1.7
Arg-76	−58.0	−4.20	−64.4	−35.2	6.4	−6.8
Gly-77	−62.5	−36.3	−63.4	−45.1	0.9	8.9
Ile-78	−64.6	−46.0	−57.6	−46.3	−7.0	0.3
Leu-79	−65.1	−17.3	−69.5	−20.3	4.4	2.9
Arg-80	−97.3	−11.6	−95.9	−6.4	−1.4	−5.2
Asn-81	−93.5	119.8	−92.7	127.3	−0.9	−7.5

the remainder of the α-helix are maintained. In the mutant structure a solvent molecule is observed 3.9 Å from the carbonyl oxygen of Asn-68, suggesting a weak hydrogen-bonding interaction.

The (ϕ, ψ) angles of the residues within the long helix are listed in Table IV. The values observed for the proline (ϕ = −59.6°, ψ = −34.6°) agree very well with the average value (ϕ = −61°, ψ = −35°) for prolines in other protein structures (MacArthur and Thornton, 1991). At the site of the substitution, the addition of the pyrrolidine ring necessitates virtually no change in ϕ. The bigger change ($\Delta\psi$ = 14°) is in the successive peptide. Not surprisingly, the largest changes in (ϕ, ψ) are for the peptide between residues 68 and 69, which is in the previous turn of the α-helix and for which the hydrogen bond to the amide of residue 72 is disrupted.

DISCUSSION

The most striking result of the present study is the finding that proline residues can be substituted at several positions within the long interdomain α-helix of T4 lysozyme with only modest effects on catalytic activity. The proteins are destabilized relative to wild-type, but still fold and behave essentially normally. Attempts were made to substitute prolines at four sites and in three cases a functional protein was obtained. It is not as if there is one particular site at which a proline can be accepted. Rather, the data suggest that it may be possible to introduce prolines at additional sites within the interdomain helix, if not at many other sites in the protein as well.

The decrease in stability observed for the two mutants D72P and Q69P is very comparable with that found for temperature-sensitive mutants of T4 lysozyme such as R96H, T157I, and A98V identified by the random genetic screen of Streisinger *et al.*, 1961; Grütter *et al.*, 1969, 1987; Weaver *et al.*, 1989; Dao-pin *et al.*, 1991a). The mutant A74P is, however, less stable than any of these previously described variants. The decrease in stability associated with the proline substitutions seems to be associated to some extent with the inaccessibility of the residue to solvent but the correlation is not perfect. Val-71 is largely buried and a proline replacement at this site did not yield a functional protein. Ala-74 is also largely solvent inaccessible. In addition the proline substitution disrupts the α-helical hydrogen bond between residues 70 and 74 which, in turn, is likely to misalign and perhaps weaken the very strong salt bridge between Asp-70 and His-31 (Anderson *et al.*, 1990). In this case protein was obtained although with substantially reduced stability. O'Neil and DeGrado (1990) found that the energy cost of an alanine to proline replacement within a dimeric α-helical model peptide was 3.4 kcal/mol. Also Yun *et al.* (1991) found, by free energy simulations, exactly the same value for an alanine to proline replacement within a short polyalanine helix. These values are roughly comparable with those found here (average values of 3.2, 2.7, and 5.2 kcal/mol) but there is no reason to expect close agreement since the context is different in every case.

In the case of A72P, for which the crystal structure is available, the pyrrolidine ring is seen to introduce three unfavorable contacts with neighboring atoms (2.93 Å between C^δ and the peptide nitrogen of Val-71; 3.04 Å between C^δ and the peptide nitrogen of Val-71; 3.00 Å between C^δ and the carbonyl oxygen of Asn-68). Each of these contacts corresponds to an unfavorable van der Waals interaction energy in the range 1–2 kcal/mol (Levitt, 1974). These steric clashes are, presumably, a major factor in the destabilization of the mutant structure relative to wild-type.

The results provide further evidence that protein structures are adaptable and can compensate for amino acid substitutions at many sites (Matthews, 1987; Sondek and Shortle, 1990). It also illustrates the redundancy that is present in the amino acid sequence of a protein. Not every amino acid in the linear sequence is necessary for folding (Reidhaar-Olson and Sauer, 1988; Zhang *et al.*, 1991). Amino acid substitutions that are expected to distort and destabilize the folded structure of the protein, and disrupt a major α-helix that might be a key folding intermediate, do not prevent the formation of a folded functional protein.

Functional proteins were obtained with prolines substituted at positions 69, 72, and 74. Residues 69 and 72 are located in successive turns on the same side of the helix. Residue 74, however, is on the *opposite* side of the α-helix. Therefore one cannot argue that prolines are only accommodated on one side of the α-helix such that each substitution bends or distorts the α-helix in the same direction.

The substitution of a proline in an α-helix that is part of a protein is not the same as a proline substitution of an isolated

α-helix. In the absence of any constraints, a substituted proline is likely to completely disrupt the helix (Strehlow *et al.*, 1991) or to introduce a bend of up to 45° (Barlow and Thornton, 1988; Karle *et al.*, 1991). The structural consequences of a proline introduced into an α-helix in a protein, however, will be modulated by interactions between the α-helix and the rest of the protein. Prolines in α-helices in known protein structures are associated with bend angles that average 26 ± 5° (Barlow and Thornton, 1988), but this is not to say that an introduced proline will cause a change of this magnitude. The determination of the structure of D72P provides an example of the structural compromise that can result. The presence of the proline increases the bend angle of the α-helix, but only by 5.5°. Thus, it is not to be expected that each individual proline substitution will cause a major rearrangement of the two domains in T4 lysozyme. Nevertheless it can be anticipated that the different substitutions will cause distinct changes in the alignment of one domain relative to the other. This suggests that the precise shape of the active-site cleft in the resting enzyme is not critical for catalysis. At the same time it should be noted that the key catalytic residues are at the base of the active site cleft, and the structural changes in this region may therefore be relatively small (≤0.5 Å). The residues in T4 lysozyme that are presumed to contact an extended oligosaccharide substrate are indicated in Fig. 4*a*.

It should be noted that the crystals that were obtained for the replacement Asp-72 → Pro are isomorphous with wild-type. It is possible that the formation of such isomorphous crystals constrains the structure of D72P to be more similar to WT* than is the case in solution. The differences between the structures of D72P and WT* in solution may therefore be larger than those seen in Figs. 3 and 4. The mutants Q69P and A74P give crystal forms that are not isomorphous with wild-type. Hopefully, the structure analysis of these crystals will give a better overall impression of the structural changes that are induced by the proline substitutions.

The first determination of the structure of a temperature-sensitive mutant lysozyme (Grütter *et al.*, 1979) showed that relatively large changes in stability were accompanied by minimal changes in structure, these being localized to the immediate vicinity of the substituted amino acid. The ability of T4 lysozyme to accommodate destabilizing mutations in this manner has subsequently been seen on a number of occasions (Matthews, 1987) and is further exemplified by the present study.

Acknowledgments—We thank Drs. Walt Baase, Larry Weaver, Eric Anderson, and Elisabeth Eriksson for advice and help with, respectively, thermodynamic measurements, data collection, NMR measurements, and crystallographic refinement. Joan Wozniak and Sheila Pepiot also provided advice in protein purification, and D. W. Heinz and X-J. "Cai" Zhang helped with data analysis. We also thank Dr. J. A. Schellman for facilities.

REFERENCES

Alber, T., and Matthews, B. W. (1987) *Methods Enzymol.* **154**, 511–533
Alber, T., Bell, J. A., Dao-pin, S., Nicholson, H., Wozniak, J., Cook, S., and Matthews, B. W. (1988) *Science* **239**, 631–635
Anderson, D. E., Becktel, W. J., and Dahlquist, F. W. (1990) *Biochemistry* **29**, 2403–2408
Anderson, W. F., Grutter, M. G., Remington, S. J., Weaver, L. H., and Matthews, B. W. (1981) *J. Mol. Biol.* **147**, 523–543
Barlow, D. J., and Thornton, J. M. (1988) *J. Mol. Biol.* **201**, 601–619
Becktel, W. J., and Schellman, J. A. (1987) *Biopolymers* **26**, 1859–1877
Bell, J. A., Wilson, K. P., Zhang, X-J., Faber, H. R., Nicholson, H., and Matthews, B. W. (1991) *Proteins Struct. Funct. Genet.* **10**, 10–21
Consler, T. G., Tsolas, O., and Kaback, H. R. (1991) *Biochemistry* **30**, 1291–1298
Dao-pin, S., Baase, W. A., and Matthews, B. W. (1990) *Proteins Struct. Funct. Genet.* **7**, 198–204
Dao-pin, S., Alber, T., Baase, W. A., Wozniak, J. A., and Matthews, B. W. (1991a) *J. Mol. Biol.* **221**, 647–667
Dao-pin, S., Soderlind, E., Baase, W. A., Wozniak, J. A., Sauer, U., and Matthews, B. W. (1991b) *J. Mol. Biol.*, **221**, 873–887
Elwell, M., and Schellman, J. (1975) *Biochim. Biophys. Acta* **386**, 309–323
Faber, H. R., and Matthews, B. W. (1990) *Nature* **348**, 263–266
Grütter, M. G., Hawkes, R. B., and Matthews, B. W. (1979) *Nature* **277**, 667–669
Grütter, M. G., Gray, T. M., Weaver, L. H., Alber, T., Wilson, K., and Matthews, B. W. (1987) *J. Mol. Biol.* **197**, 315–329
Hawkes, R., Grütter, M. G., and Schellman, J. (1984) *J. Mol. Biol.* **175**, 195–212
Jenks, W. P. (1969) *Catalysis in Chemistry and Enzymology*, McGraw-Hill, New York
Jones, T. A. (1978) *J. Appl. Crystallogr.* **11**, 268–272
Karle, I. L., Flippen-Anderson, J. L., Agarwalla, S., and Balaram, P. (1991) *Proc. Natl. Acad. Sci. U. S. A.* **88**, 5307–5311
Knowles, J. (1991) *Nature* **350**, 121–124
Kunkel, T. A. (1985) *Proc. Natl. Acad. Sci. U. S. A.* **82**, 488–492
Kunkel, T. A., Roberts, J. D., and Zakour, R. A. (1987) *Methods Enzymol.* **154**, 367–382
Levitt, M. (1974) *J. Mol. Biol.* **82**, 393–420
MacArthur, M. W., and Thornton, J. M. (1991) *J. Mol. Biol.* **218**, 397–412
Matsumura, M., and Matthews, B. W. (1989) *Science* **243**, 792–794
Matthews, B. W. (1987) *Biochemistry* **26**, 6885–6888
Matthews, B. W., Nicholson, H., and Becktel, W. J. (1987) *Proc. Natl. Acad. Sci. U. S. A.* **84**, 6663–6667
Muchmore, D. C., McIntosh, L. P., Russell, C. B., Anderson, D. E., and Dahlquist, F. W. (1989) *Methods Enzymol.* **177**, 44–73
Nishikawa, K., Ooi, T., Isogai, Y., and Saito, N. (1972) *J. Physiol. Soc. Jpn.* **32**, 1331–1337
O'Neil, K., and DeGrado, W. L. (1990) *Science* **250**, 646–651
Pjura, P., Matsumura, M., Wozniak, J., and Matthews, B. W. (1990) *Biochemistry* **29**, 2592–2598
Poteete, A. R., Dao-Pin, S., Nicholson, H., and Matthews, B. W. (1991) *Biochemistry* **30**, 1425–1432
Reidhaar-Olson, J. F., and Sauer, R. T. (1988) *Science* **241**, 53–57
Sondek, J., and Shortle, B. (1990) *Proteins Struct. Funct. Genet.* **7**, 299–305
Storm, D. R., and Koshland, D. E., Jr. (1970) *Proc. Natl. Acad. Sci. U. S. A.* **66**, 645
Strehlow, K. G., Robertson, A. D., and Baldwin, R. L. (1991) *Biochemistry* **30**, 5810–5814
Streisinger, G., Mukai, F., Dreyer, W. J., Miller, B., and Horiuchi, S. (1961) *Cold Spring Harbor Symp. Quant. Biol.* **26**, 25–30
Tronrud, D. E., Ten Eyck, L. F., and Matthews, B. W. (1987) *Acta Crystallogr.* **A 43**, 489–503
Tsugita, A., Inouye, M., Terzaghi, E., and Streisinger, G. (1968) *J. Biol. Chem.* **243**, 391–397
Weaver, L. H., and Matthews, B. W. (1987) *J. Mol. Biol.* **193**, 189–199
Weaver, L. H., Gray, T. M., Grütter, M. G., Anderson, D. E., Wozniak, J. A., Dahlquist, F. W., and Matthews, B. W. (1989) *Biochemistry* **28**, 3793–3797
Wetzel, R., Perry, L. J., Baase, W. A., and Becktel, W. J. (1988) *Proc. Natl. Acad. Sci. U. S. A.* **85**, 401–405
Yun, R. H., Anderson, A., and Hermans, J. (1991) *Proteins Struct. Funct. Genet.* **10**, 219–228
Zhang, X-J., Baase, W. A., and Matthews, B. W. (1991) *Biochemistry* **30**, 2012–2017

The Journal of Biological Chemistry 1992

Appendix C

JOURNAL OF
MEDICINAL
CHEMISTRY

Volume 33, Number 3 March 1990

Special Topic

Molecular Modeling Software and Methods for Medicinal Chemistry†

N. Claude Cohen,‡ Jeffrey M. Blaney,‡,§ Christine Humblet,‖ Peter Gund,⊥ and David C. Barry#

Ciba-Geigy Ltd. Pharmaceutical Division, Basel, Switzerland, du Pont de Nemours & Company, Wilmington, Delaware 19898, Parke-Davis Pharmaceutical Research Division, Ann Arbor, Michigan 48105, Merck Sharp & Dohme Research Laboratories, Rahway, New Jersey 07065, and ICI Pharmaceuticals Division, Alderley Park, Cheshire, U.K. *Received May 11, 1989*

I. Introduction

Molecular modeling has become a well-established research area during the last decade due to advances in computer hardware and software that have brought high-performance computing and graphics within the reach of most academic and industrial laboratories. A growing number of journals now focus on molecular modeling: *Journal of Computational Chemistry, Computers in Chemistry, Journal of Computer-Aided Molecular Design, Journal of Molecular Graphics, Molecular Simulations,* and *Tetrahedron Computer Methodology.* Several recent texts and reviews describe progress in molecular modeling research and applications.[1-7]

This review is intended to provide medicinal chemists with introductory material related to available molecular modeling software and methods. A particular emphasis is given to current software that integrates multiple methods, including graphic and computational tools, and focuses on systems familiar to the committee.

It is important to realize what is really meant by "computer-assisted drug design". Molecular modeling systems provide powerful tools for *building, visualizing, analyzing,* and *storing* models of complex molecular systems that can help interpret structure–activity relationships. The critical problem of *molecular design*—what structure do we build, model, and possibly synthesize?—is not answered by current methods and is left up to the creativity of the medicinal chemist. The goal of molecular modeling should not be limited only to providing insight, but it should also help to suggest new experiments, i.e., new structures tailored to have the desired biological activity. Molecular modeling cannot yet produce quantitative predictions of activity except in very special cases, but it can provide valuable qualitative guidelines that help design new lead structures. The result of a successful modeling study is therefore usually one or more candidate structures predicted to fulfill particular criteria described in a molecular model, i.e., a pharmacophore. The synthesis and biological evaluation of these target structures can be used to test and iteratively refine the model.

"Direct" and "indirect" design are the two major modeling strategies currently used in the conception of new drugs. In the first approach the three-dimensional features of a known receptor site are directly considered, and in the latter the design is based on the comparative analysis of the structural features of known active and inactive molecules that are interpreted in terms of complementarity with a hypothetical receptor site model (Figure 1). Spe-

† This is the second of three Special Topics on the subject of Molecular Modeling in Drug Design commissioned by the Committe on Medicinal Chemistry of IUPAC (Toplis, J. G. *J. Med. Chem.* 1988, *31*, 2229). The first article, *Guidelines for Publications in Molecular Modeling Related to Medicinal Chemistry*, appeared in an earlier issue (Gund, P.; Barry, D. C.; Blaney, J. M.; Cohen, N. C. *J. Med. Chem.* 1988, *31*, 2230). A third article on Molecular Modeling Hardware is in preparation.

‡ Ciba-Geigy Ltd. Pharmaceutical Division.
§ du Pont de Nemours & Company.
‖ Parke-Davis Pharmaceutical Research Division.
⊥ Merck Sharp & Dohme Research Laboratories.
ICI Pharmaceuticals Division.

(1) Cohen, N. C. *Drugs Future* 1985, *10*, 311.
(2) Cohen, N. C. In *Advances in Drug Research*; Testa, B., Ed.; Academic Press: 1985; Vol. 14, p 41.
(3) Ripka, W. C. *Nature* 1986, *21*, 93.
(4) Burgen, A. S. V.; Roberts, G. C. K.; Tute, M. S. *Molecular Graphics and Drug Design*; Topics in Molecular Pharmacology; Vol. 3; Elsevier: Amsterdam, 1986;
(5) Gund, P.; Halgren, T. A.; Smith, G. M. *Annu. Rep. Med. Chem.* 1987, *22*, 269.
(6) Sheridan, R. P.; Venkataraghavan, R. *Acc. Chem. Res.* 1987, *20*, 322.
(7) Dean, P. M. *Molecular foundations of drug-receptor interaction*; Cambridge University Press: Cambridge, 1987.

0022-2623/90/1833-0883$02.50/0

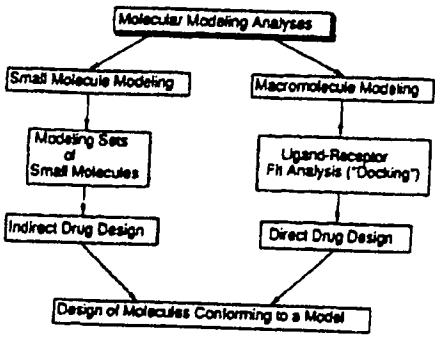


Figure 1.

cialized molecular modeling systems have been developed to analyze either the interaction of a prototype molecule with a known receptor site or the ability of a given compound to mimic the three-dimensional stereochemical features of known active compounds. Both approaches attempt to optimize receptor fit for selectivity and binding affinity while qualitatively considering other critical factors (log *P*, solubility, metabolic stability, etc.)

Most molecular modeling systems strive to provide the same basic set of features: visualization and manipulation of three-dimensional molecular models including rotatable bonds, structure building, molecular mechanics and/or dynamics, conformational analysis, electronic properties, molecular surface displays, and the calculation of various physical properties.

II. Interactive Graphics Display and Manipulation

A large range of graphics workstations are available to meet the needs of modeling applications ranging from simple, small molecule to complex macromolecules. For small molecules basic, inexpensive systems may be adequate (e.g. a Macintosh II can handle up to a couple hundred atoms in real time; real time means that the molecular model rotates and translates smoothly under interactive control). Current personal computer (PC) molecular modeling software have been reviewed recently.[175,187] The sheer size of macromolecules requires sophisticated graphics software and hardware to provide real-time, interactive response along with selective display and manipulation.[8] Current state-of-the-art systems are capable of simultaneously handling up to 20 or more molecules with up to about 20 000 atoms and thousands of molecular surface points in real time with depth-cued color and time-sliced stereo. Each molecule should be able to be individually labeled, color-coded, and controlled in three dimensions, while simultaneously monitoring inter and/or intramolecular distances and adjusting multiple contiguous or noncontinguous dihedral angles. Dials, joysticks, and a mouse, or an excellent new interactive device called "Spaceball",[9] which simultaneously control all six degrees of rotational and translational freedom with a single hand, are used to translate and rotate molecules and to rotate bonds. Typical operations are activated by first pointing to a menu and next to atoms and bonds, either with a stylus or a "mouse" to calculate, for example, distances and angles (dihedral or valence). Most systems continually update this information as the geometries are modified. The latest graphics workstations have very fast processors that do complete bump-checking (checking for contacts closer than van der Waals) and even molecular mechanics and dynamics energy calculations in real time (for small molecules up to about the size of a decapeptide). Selective control of which molecules or portions of molecules are displayed and which molecules, distances, and dihedral angles are active requires a powerful command language along with interactive "picking" of atoms and bonds with a mouse or stylus.

The trend in recent molecular modeling software design has been to exploit the powerful new windowing and computational power of the new generation of graphics workstations. This has resulted in an emphasis on menu-driven systems, which are intuitive and easy to learn, but sacrifice generality and completeness if not carefully implemented. Menu designs provide the most basic commands, but the complex syntax required by nearly all the current systems' command languages makes specifying functions not found on the menus cumbersome, if not impossible, for the nonspecialist. Hopefully, continued software design efforts will create improved menu systems and realize the need for simple, English-like command language syntax to supplement features not easily implemented in menus. The new design trend has also focused on integrating computational chemistry (e.g. molecular mechanics and dynamics) with graphics display, but much of the effort has been devoted to computations, at the expense of neglecting important features and a good user interface for interactive graphics pioneered in previous generations of graphics-only modeling systems. Despite the impressive computational performance of the new workstations, even the most sophisticated techniques provide only rough, qualitative guidance for most medicinal chemistry applications. Good interactive graphics with a well-designed user interface maximizes the performance of the most critical part of the system—the chemist.

Raster graphics has recently become the dominant technology in interactive molecular modeling, replacing the older calligraphic or vector display systems. Although raster displays have apparent advantage in providing beautiful "realistic" color solid shaded images, these images cannot be updated fast enough (with transparency and clipping) for real-time modeling yet, so vector and dot images (on raster displays) still provide the best approach for high-performance molecular modeling. Vector (bonds) and dot (molecular surface) images have the tremendous advantage of providing full transparency and clipping while displaying a complex, color-coded molecular surface and bonds, which are essential for studying interactions deep inside a macromolecular binding site or comparing several small molecules.[8] Time-sliced stereo, where the left and right eye views are alternately displayed approximately every $^1/_{30}$ s and viewed through a mechanical shutter or liquid crystal glasses synchronized to the display, provides a very convincing three-dimensional illusion and is extremely helpful for modeling complex interactions. A recent major improvement in stereo viewing is to place a liquid crystal screen over the entire graphics screen, allowing the user(s) to wear circularly polarized plastic glasses.

The simultaneous development of real-time interactive color graphics[8] and Connolly's molecular surface program[10] in 1980 revolutionized molecular modeling. Col r-coded surfaces provide qualitative displays of hydrophobic and hydrophilic regions, neutral and charged groups, electr -

(8) Langridge, R.; Ferrin, T. E.; Kuntz, I. D.; Connolly, M. L. *Science* 1981, *211*, 661.
(9) Spatial Systems Pty Ltd., PO Box 452, 55 Lavender St., Milsons Point, NSW 2061, Australia.
(10) Connolly, M. L. *Science* 1983, *221*, 709.

static potential, and mobility (based on X-ray crystallographic refinement or molecular dynamics simulation). Color-coded molecular surfaces therefore simultaneously display the main features critical to receptor binding: shape, charge, and hydrophobicity. Hydrophobic color coding was originally done simply by coloring all surface points associated with carbon "hydrophobic" (e.g. red) and all nitrogen and oxygen surface points "hydrophilic" (e.g. blue); an improved approach[11] includes "neutral" surface (e.g. yellow) for sulfur, α-carbons of amino acids, the carbon between the imidazole nitrogens in histidine, and carbonyl carbon. Molecular surfaces can also be color coded by a so-called "hydrophobic potential", based on fragment hydrophobicity values and a simple empirical function analogous to the classical formula for electrostatic potential.[12,13] Electrostatic potential molecular surfaces[14] are calculated using quantum mechanically derived partial atomic charges for each atom.[15] The potential is usually calculated one probe sphere radius above the molecular surface to give a qualitative view of what an incoming ligand "sees" as it approaches the macromolecule. The surface is color coded by the value of the electrostatic potential at each point. The electrostatic potential gradient or electric field can also be displayed graphically using short vectors.[16] Similar representations can also be envisaged for any other potential or field such as, for example, the molecular mechanics potential experienced by different chemical probes.[139]

Connolly's program[10] implemented Richard's definition[17] of molecular surface by rolling a probe sphere (usually 1.4-Å radius, the effective radius of water molecule) over the surface of the molecule, resulting in a smooth surface that represents the surface accessible to a water molecule, including internal cavities. Langridge's UCSF group[18] and Pearle and Honneger[19] independently developed van der Waals dot surface programs that are much faster than Connolly's molecular surface program, although they are not as effective at eliminating buried surface and produce a more complicated surface display for macromolecules. Both types of surface are available in most modeling systems. Connolly also developed an analytical method for calculating molecular surface,[20] which provides nearly exact values for the surface area and volume[21] enclosed by a surface along with spectacular shaded raster graphics images,[22] which gives a much different impression of a surface than the conventional CPK-like raster surfaces.[23] Barry introduced the very useful "extra radius" surface,[24] where the surface is calculated one van der Waals radius beyond the normal surface, collapsing the surface of a binding site onto the vector model of its ligand and eliminating the need for displaying the ligand's surface. This simple graphics trick makes it much easier to visualize the "docking" of a ligand into a binding site. For example, chymotrypsin's specificity for aromatic amino acid side chains is not immediately apparent from a conventional molecular surface of its active site, while the "extra radius" surface reveals an almost perfectly planar rocket that is obviously complementary to an aromatic ring. The "extra radius" surface can also be color coded by hydrophobicity or electrostatic potential.

III. Small Molecule Modeling

(a) Structure Building. Every system should provide means allowing one to construct accurate three-dimensional models of organic molecules. One of the simplest and most reliable ways is to use libraries of typical organic fragments and the Cambridge X-ray Crystallographic Data Base,[25] which contains about 50 000 structures. A molecule is constructed by assembling preexisting fragments, followed by successive adjustments of the current structure, which allows the user full control over building a reasonable starting conformation with the desired stereochemistry. Several common building functions were involved in these operations: make-bond, break-bond, fuse-rings, delete-atom, add-atom, add-hydrogens, invert chiral center, etc. They are combined with continuous refinements f the geometry of the current structure using molecular mechanics.

Most systems have facilities allowing one to draw chemical structures as a two-dimensional sketch describing the atom types (element and hybridization) and connectivity (what's bonded to what), along with some method of specifying stereochemistry (up/down, *R*/*S*, etc.). While in principle a simple and intuitive approach, it has proven very challenging to design robust methods to convert the initial two-dimensional information into reasonable low energy conformations. Most of these approaches are molecular mechanics, but often become trapped quickly in poor local minima during the conversion from two into three dimensions. Distance geometry combined with molecular mechanics[26,27] usually provides superior results to molecular mechanics alone. Very few systems are able to handle the conformational multiplicity of cyclic moieties in a fully automatic manner.[72,150] Pearlman[28] recently introduced CONCORD, an elegant method for rapidly generating good quality three-dimensional structures directly from a SMILES[29] code (a simple alphanumeric language for encoding organic structures). CONCORD is currently the best available method for generating small-molecule three-dimensional structures interactively, due to its ease of use, speed, and the quality of the resulting structure. It has the advantage of being able to produce a good quality structure for most organic compounds, including those with complex heteroatom functional groups and ring systems, without the need for developing molecular mechanics parameters. However, CONCORD

(11) Recanatini, M.; Klein, T.; Yang, C.; McClarin, J.; Langridge, R.; Hansch, C. *Mol. Pharmacol.* 1986, *29*, 436.
(12) Fauchere, J. L.; Quarendon, P.; Kaetterer, L. *J. Mol. Graphics* 1988, *6*, 203.
(13) Furet, P.; Sele, A.; Cohen, N. C. *J. Mol. Graphics* 1988, *6*, 182.
(14) Weiner, P. K.; Langridge, R.; Blaney, J. M.; Schaefer, R.; Kollman, P. A. *Proc. Natl. Acad. Sci. U.S.A.* 1982, *79*, 3754.
(15) Singh, U. C.; Kollman, P. A. *J. Comput. Chem.* 1984, *5*, 129.
(16) Getzoff, E. D.; Tainer, J. A.; Weiner, P. K.; Kollman, P. A.; Richardson, J. S.; Richardson, D. C. *Nature* 1983, *306*, 287.
(17) Richards, F. M. *Annu. Rev. Biophys. Bioeng.* 1977, *6*, 151.
(18) Bash, P. A.; Pattabiraman, N.; Huang, C.; Ferrin, T. E.; Langridge, R. *Science* 1983, *222*, 1325.
(19) Pearl, L. H.; Honegger, A. *J. Mol. Graphics* 1983, *1*, 9.
(20) Connolly, M. L. *J. Appl. Crystallogr.* 1983, *16*, 548.
(21) Connolly, M. L. *J. Am. Chem. Soc.* 1985, *107*, 1118.
(22) Connolly, M. L. *J. Mol. Graphics* 1985, *3*, 19.
(23) Feldman, R. J.; Bing, D. H.; Furie, B. C.; Furie, B. *Proc. Natl. Acad. Sci. U.S.A.* 1978, *75*, 5409.
(24) Barry, C. D. Unpublished results.
(25) Allen, F. H.; Bellard, S.; Brice, M. D.; Cartwright, B. A.; Doubleday, A.; Higgs, H.; Hummelink, T.; Hummelink-Peters, B. G.; Kennard, O.; Motherwell, W. D. S.; Rodgers, J. R.; Watson, D. G. *Acta Crystallogr.* 1979, *B35*, 2331.
(26) Wenger, J. C.; Smith, D. H. *J. Chem. Inf. Comput. Sci.* 1982, *22*, 29.
(27) Weiner, P. K.; Profeta, S., Jr.; Wipff, G.; Havel, T.; Kuntz, I. D.; Langridge, R.; Kollman, P. A. *Tetrahedron* 1983, *39*, 1113.
(28) Rusinko, A., III; Skell, J. M.; Balducci, R.; Pearlman, R. S. *CONCORD*, University of Texas at Austin; distributed by Tripos Associates, St. Louis, MO, 1987.
(29) Weininger, D. *J. Chem. Inf. Comput. Sci.* 1988, *28*, 31.

generates only a single conformer and cannot be used for conformational sampling. CONCORD has also been used to generate three-dimensional structures from two-dimensional structures stored in large industrial databases to provide conformations for newly developing three-dimensional search techniques.[30]

Many popular file formats for storing three-dimensional coordinates are in use (Brookhaven Protein Data Bank, Cambridge, Molecular Design's MOLFILE, CHEM-X' CSSR, etc.), but unfortunately there is no accepted convention or standard. The best current solution, used by more and more modeling systems to provide compatibility with other software, is to include facilities to read and write most or all of the popular formats, while making it easy for the user to add new formats. A standard molecule file format has been proposed.[160]

Molecular modeling studies result in a proliferation of files containing different results from different theoretical and experimental methods. Keeping track of all this data for several different projects can easily become a bookkeeping nightmare. Several current systems provide simple databases for storing and retrieving the results generated. A more general solution is provided by THOR,[32] an elegant chemical database system based on SMILES[29] codes. Martin et al.[33] described the use of THOR for molecular modeling.

(b) Molecular Mechanics. Molecular mechanics methods[34,35] are based on a pragmatic view of the molecular structure that is considered as a set of balls and springs with series of potential energy functions expressing the molecular force field as a sum of these functions. A typical energy equation is as follows:

$$E_{total} = E_{stretching} + E_{bending} + E_{dihedral} + E_{van\ der\ Waals} + E_{electrostatic} + E_{hydrogen\ bond}$$

Each of the individual energy terms have preferential equilibrium positions (bond lengths, bond angles, dihedral angles, van der Waals interaction distances, etc.) and force constants that are either experimentally known or theoretically estimated and used to associate energetic penalties with each individual deviation. A "Force Field" therefore consists of a set of analytical energy functions and their associated sets of numerical parameters. The total energy of a given molecule can be the sum of several thousands of individual contributions. Force field development remains a major problem for the large variety of complex functional groups encountered in medicinal chemistry, which is further complicated by the fact that not all force fields are readily transferable from one package to another. The most extensively tested force fields are MM2[34] (hydrocarbons plus a limited selection of simple heteroatom functional groups), AMBER[36-38] and CHARMM[39] (peptides and nucleic acids), and ECEPP[173,174] (peptid s). MM2 is the current standard for small-molecule work, but is a poor choice f r macromolecules. AMBER and CHARMM force fields are similar and are the standard for macromolecules, but give only qualitative results on small molecules. Hybrid force fields, such as the AMBER all-atom force field,[38] are usually used for calculations involving small-molecule–macromolecule interactions. Molecules that contain functional groups not parameterized by the above force fields require the estimation of new parameters specific for each new bond, bond angle, or dihedral angle type.[40] Most of the major software systems provide facilities for automatically assigning the appropriate atom types and parameters, but there is considerable variation in the quality and quantity of the parameters available. It is always prudent to calibrate unfamiliar software with some well-known test cases. Biosym[41] has formed an industrial consortium to systematically develop and test force field parameters. Assuming that all the necessary parameters are available for a given molecule, relative total strain energies can be calculated for estimating rotation or inversion barriers, preferred conformations, the energy required to achieve a specific conformation, etc. Except for special cases (e.g. estimating the enthalpy of formation of a hydrocarbon) the absolute calculated energy is of little value—*relative* energies between different conformers or isomers are important. The texts by Buckert and Allinger[34] and Clark[42] provide an excellent description of molecular mechanics and its applications.

Molecular mechanics energy minimization involves successive iterative computations, where an initial conformation is submitted to full geometry optimization. All parameters defining the geometry of the system are modified by small increments until the overall structural energy reaches a local minimum. The goal is to reach a local minimum on the potential surface within the minimum amount of time. The more sophisticated methods use the first and occasionally the second derivatives of the energy function for guiding the minimization. No method can guarantee finding the absolute lowest energy structure—the global minimum. Energy minimization will stop at the first local minimization encountered, with ut realizing that much deeper, more stable minima may be accessible. The problem is analogous to a ball rolling downhill, which stops in the first valley it finds and is unable to climb the next hill which may lead to a deeper valley. Molecular dynamics is able to climb small barriers (the barrier height depends on the temperature of the dynamics simulation) and is therefore much more efficient at locating deep local minima than simple minimizati n; short dynamics runs are now commonly used for minimization. Systematic search,[43,44] which increments all rotatable bonds in turn to explore the complete conformation space of the molecule, distance geometry[45,46] and

(30) Brint, A. T.; Willett, P. *J. Mol. Graphics* **1987**, *5*, 49.
(31) Chem-X, developed and distributed by Chemical Design Ltd., Oxford, England.
(32) Weininger, D.; Weininger, A. *THOR—THeaurus ORiented chemical database, version 3.54*; Daylight Chemical Information Systems: Claremont, CA 91711, 1989.
(33) Martin, Y. C.; Danaher, E. B.; May, C. S.; Weininger, D. *J. Comput.-Aided Mol. Des.* **1988**, *2*, 15.
(34) Buckert, U.; Allinger, N. L. *Molecular Mechanics*; American Chemical Society: Washington, DC, 1982.
(35) Osawa, E.; Musso, H. In *Topics in Stereochemistry*; Allinger, N. L., Eliel, E. L., Wilen, S. H., Eds.; Wiley: New York, 1982; Vol. 13, p 117.
(36) Weiner, P. K.; Kollman, P. A. *J. Comput. Chem.* **1981**, *2*, 287.
(37) Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S., Jr.; Weiner, P. *J. Am. Chem. Soc.* **1984**, *106*, 765.
(38) Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; Case, D. A. *J. Comp. Chem.* **1986**, *7*, 230.
(39) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comp. Chem.* **1983**, *4*, 187.
(40) Hopfinger, A. J. *J. Comp. Chem.* **1984**, *5*, 486.
(41) Biosym Technologies, Inc., 10065 Barnes Canyon Rd., San Diego, CA 92121.
(42) Clark, T. *A Handbook of Computational Chemistry*; John Wiley and Sons: New York, 1985.
(43) Dammkoehler, R. A.; Darasek, S. F.; Berkely Shands, E. F. *J. Comput.-Aided Mol. Des.* **1989**, *3*, 3.
(44) Motoc, I.; Dammkoehler, R. A.; Marshall, G. R. In *Mathematical and Computational Concepts in Chemistry*; Trinajstic, N., Ed.; Horwood, Ltd.: Chichester, 1986; p 222.

other random sampling approaches attempt to locate the global minimum through thorough exploration of the allowed conformations, while the ellipsoid method[47,48] and an extension of distance geometry called energy embedding[49] can accomplish near global optimization in some cases.

Energy minimization can proceed either in internal coordinates (the variables explicitly considered are the bond lengths, bond angles, and dihedral angles) or, as is more often the case, in Cartesian coordinates (each atom is characterized with x, y, and z coordinates, and the atom moves with small increments along these axes). An advantage of minimizing in internal coordinates is that cooperative movements of several atoms or groups are well simulated in such treatments; moreover since the degrees of freedom of the chemical structures are natural, the risk that the molecules are trapped in a false minima is greatly reduced.

(c) Molecular Dynamics. In the last 10 years the static views of molecules have been considerably enlarged to include new perspectives introduced by molecular dynamics.[50,51] X-ray crystal structures represent a time-averaged structure of a continuously moving system, while molecular dynamics simulates the actual, instantaneous motion of the system. Each atom is treated as a particle responding to Newton's equations of motion: successive integrations of these equations lead to the trajectory of the atom over time in the form of a list of positions and velocities. Analyses are made through periods of typically 1–100 ps (many interesting motions are fully developed within 100 ps or less).

The motions of the atoms and chemical groups obtained by these simulations reveal subtle underlying molecular machinery and make it possible to understand phenomena that cannot be explained by the static view. Over short periods of time (e.g. a fraction of a picosecond), molecular dynamics usually shows little coherence in the displacements of the atoms. The motions are frequently interrupted by collisions with neighboring groups, and each group seems to have an erratic trajectory. Over longer periods of time, coherent and collective motions start to develop, revealing how some groups can fluctuate somewhat more than others.

The calculations require good computational power as well as appropriate graphical facilities. Animation consists of the viewing of consecutive conformations generated by molecular dynamics calculations. Animated display of molecular dynamics simulations is essential; dynamics simulations produce huge amounts of data that are difficult to interpret without graphics.

Moelcular dynamics is useful in order to identify preferred motions of either small molecules or proteins. Although it is not of direct utility in drug design except for "where does it spend most of its time" and as an improved energy minimization approach, dynamics gives a high information content picture of the precise behavior of the molecule considered and the way it can behave and interact with other partners. Restrained molecular dynamics[52] adds an artificial penalty function to restrain specific distances, angles, r dihedral angles. Restrained molecular dynamics and distance geometry[53,54] have been used to generate three-dimensional structures of small molecules, proteins, and nucleic acids consistent with NMR data.[55] Multiple energy minimization force fields are used in molecular dynamics methods and have been described in the literature.[176-186] Recent reviews[176,177] provide excellent description of molecular dynamics and related methods and illustrate various application approaches.

(d) Quantum Mechanics. In principle all treatments mentioned in the preceding paragraph can be made by using quantum chemical calculations. Molecular energies are calculated by using the Schroedinger equation with the Molecular Orbital (MO) formalism, which can provide greater accuracy along with the ability to model electronic effects not treated by molecular mechanics, as well as consume enormous amounts of computer time depending on the method and approximations used. Over a long period of time the Quantum Chemical Program Exchange (QCPE) group located at the University of Indiana has contributed greatly to the dissemination of a number of excellent theoretical chemistry programs to the scientific community.

The Schroedinger equation of a given molecular system can be solved either with no approximations at all (ab initio) or with the introduction of some approximations (semiempirical). Semiempirical treatments such as AM1,[56] MNDO,[57] CNDO[58,59] INDO,[60] EHT, MINDO,[61] PRDDO,[62] and PCILO[63,64] are some of the most popular semiempirical programs, whereas the GAUSSIAN[65] and HONDO[66] series are typical ab initio programs. AMPAC and MOPAC are QCPE packages that include the AM1, MNDO, and MINDO programs. Along with GAUSSIAN series, these are among the most popular programs for quantum mechanical calculations.[67]

(45) Crippen, G. M. *Distance Geometry and Conformational Calculations*; Bawden, D., Ed.; Research Studies Press (Wiley): New York, 1981.
(46) Crippen, G. M.; Havel, T. F. *Distance Geometry and Molecular Conformation*; Bawden, D., Ed.; Research Studies Press (Wiley): New York, 1988.
(47) Billeter, M.; Havel, T. F.; Wuthrich, K. *J. Comp. Chem.* 1987, *8*, 132.
(48) Billeter, M.; Havel, T. F.; Kuntz, I. D. *Biopolymers* 1987, *26*, 777.
(49) Crippen, G. M. *J. Phys. Chem.* 1987, *91*, 6341.
(50) Karplus, M.; McCammon, J. A. *Annu. Rev. Biochem.* 1983, *52*, 263.
(51) McCammon, J. A.; Harvey, S. C. *Dynamics of Proteins and Nucleic Acids*; Cambridge University Press: Cambridge, 1987.
(52) Clore, G. M.; Nilges, M.; Brunger, A. T.; Karplus, M.; Gronenborn, A. M. *FEBS Lett.* 1987, *213*, 269.
(53) Havel, T.; Wuthrich, K. *Bull. Math. Biol.* 1984, *46*, 673.
(54) Braun, W.; Go, N. *J. Mol. Biol.* 1985, *186*, 611.
(55) Wuthrich, K. *NMR of Proteins and Nucleic Acids*; John Wiley and Sons: New York, 1986.
(56) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* 1985, *107*, 3902.
(57) Dewar, M. J. S.; Thiel, W. *J. Am. Chem. Soc.* 1977, *99*, 4899.
(58) Pople, J. A.; Segal, G. A. *J. Chem. Phys.* 1965, *43*, S136.
(59) Pople, J. A.; Segal, G. A. *J. Chem. Phys.* 1966, *44*, 3289.
(60) Pople, J. A.; Beveridge, D. L.; Dobosh, P. A. *J. Chem. Phys.* 1967, *47*, 2026.
(61) Bingham, R. C.; Dewar, M. J. S.; Lo, D. H. *J. Am. Chem. Soc.* 1975, *97*, 1302.
(62) Halgren, T. A.; Kleier, D. A.; Hall, J. H., Jr.; Brown, L. D.; Lipscomb, W. N. *J. Am. Chem. Soc.* 1978, *100*, 6595.
(63) Diner, S.; Malrieu, J. P.; Claverie, P. *Theor. Chim. Acta* 1969, *13*, 1.
(64) Diner, S.; Malrieu, J. P.; Jordan, F.; Gilbert, M. *Theor. Chim. Acta* 1969, *15*, 100.
(65) Hehre, W. J.; Radom, L.; Schleyer, P. v. R.; Pople, J. A. *Ab Initio Molecular Orbital Theory*; John Wiley & Sons: New York, 1986.
(66) Dupuis, M.; Rys, J.; King, H. F. *HONDO*, Quantum Chemistry Program Exchange, Indiana University: Bloomington, 1976.
(67) Popular programs distributed by QCPE include: MOPAC (455), AM1 (506), MNDO (428), CNDO/INDO (389), EHT (358), MINDO (309), PCILO (220), GAUSSIAN82 (446), HONDO (403), AMPAC (506).

Energies can be obtained through either the "self consistent" (SCF) formalism or with "perturbation methods". The SCF method is based on a property of the Schroedinger equation which states that whatever wave function is used to calculate the electronic energy of a given system, the corresponding energy will always be greater than the true energy value. SCF treatments are based on that property as follows: starting with an initial wave function, iteratively modify it until the total energy does not decrease. Full geometry optimizations therefore require the combination of two types of minimization: one for the calculation of the energies, and one for the optimization of the geometries.

In the perturbation methods, as in PCILO approaches,[63,64] the total energy is calculated as a convergent series of terms, with each new term improving the accuracy of the previously computed energy. The approach starts from the initial two-dimensional chemical formula that is used to compute the first term of the series. In general the treatment is stopped either at the second or at the third order. An advantage of these computations is that they are relatively rapid and permit one to obtain "conformational maps" (e.g. energy contours according to the variation of two dihedral angles). The computer time necessary to calculate a map using a 30-deg increment (12 × 12 = 144 conformations) is comparable in perturbation methods to the time necessary for only one or two conformations using SCF methods.

Quantum chemical calculations can provide detailed insight into the electronic nature of the molecular structures and allow one to analyze phenomena not yet parameterized for molecular mechanics. Molecular mechanics calculations compete favorably with MO calculations for conformational analysis and can be applied to much larger molecules; however, there are a number of physical, chemical, and electronic indices that can be obtained only with quantum mechanical treatments. These methods are theoretically powerful and can be very useful, but the tremendous amount and variety of data they generate must be interpreted with care. In some treatments, particularly when it is known that different methods might not lead to the same results, it is safer to pay more attention to the variations and the trends of the molecular property analyzed rather than to consider their absolute values. A well-known example of lack of agreement of different methods is the calculation of partial atomic charges, which are required by most molecular mechanics force fields and for the calculation of molecular electrostatic potentials. Several approaches have been developed for calculating partial atomic charges in molecules.[15,68-70] Current knowlege of the strengths and weaknesses of available semiempirical and ab initio methods was recently reviewed in an excellent introductory text.[42] Richards' text[71] provides a good introduction into applications of quantum mechanical calculations for medicinal chemistry.

In practice only molecules containing less than about 50 atoms can be studied with quantum mechanical approaches. The selection of the most appropriate method depends not only on the size of the molecule but also on the type of molecular property (e.g. conformation, electronic density, electrostatic potential, frontier orbitals, etc.) that is desired. Most major molecular modeling software packages provide interfaces to popular quantum mechanical methods.

(e) Conformational Analysis. In a first approximation, only intramolecular forces are considered to calculate the conformational properties of a given molecule. However, force field treatments are not restricted to isolated molecules ("gas phase simulations"), they can be envisaged with two molecules as in "docking" analyses, or even simulate solvent molecules in the investigation of solvent effects. Since the global energy minimum is not necessarily the receptor-bound conformation, it is essential to sample a region up to several kilocalories/mole above the global minimum. Molecular mechanics approaches are commonly used for conformational analysis, but quantum mechanical methods can be used for small molecules with two to three rotatable bonds.

A multiple conformation generation function appears now in an increasing number of modeling systems, but is often restricted to the rotation of acyclic bonds. Few modeling systems are able to handle the conformational multiplicity of cyclic (monocyclic or polycyclic) systems automatically. A robust method based on conformational assembly rules has been described[72] allowing the systematic and automatic generation of possible conformations of simple or complex cyclic molecules having, for example, precise polycyclic fused, spiro and bridge-headed systems (when the size of the rings is relatively small, e.g. less than eight members for each elementary ring). Smith et al.[73] described a variation of systematic search for cyclic systems. Gerber et al.[155] developed an elegant method for the systematic generation of conformations in macrocyclic systems that is based on generic shapes approximated by Fourier harmonic representations. More general methods based on artificial intelligence techniques were proposed to generate reliable low-energy conformations of any given small molecule.[74] Efficient variations of systematic search techniques have been described by Dammkoehler et al.[43,44] and Lipton.[75] Chang et al.[128] recently described a new Monte Carlo (random) torsion search method that appears to be one of the most efficient approaches for small molecule conformational analysis. Most major molecular modeling systems include approaches, along with extensive analysis facilities (e.g. contour plots of energy as a function of two dihedral angles). Scheraga and Colleagues have developed a series of techniques in conformational searching of polypeptides (for a review, see ref 169) that include build-up procedures,[170] increase of dimensionality,[171] Monte Carlo plus minimizations,[172] and optimizati n of electrostatics.[169]

Distance geometry calculations can also be used to generate random starting conformations for conformational analysis.[26,27] Distance geometry is a general method for converting a set of distance constraints into a set f three-dimensional coordinates consistent with the constraints.[45,46] The distance constraint matrix describes the complete conformation space of a molecule by including the maximum possible distance (upper bond) between each atom pair and the minimum possible distance (lower bound). All possible conformers lie between these upper and lower distance bound—distance geometry converts this distance information into three-dimensional coordinates.

(68) Pepe, G.; Serres, B.; Laporte, D.; Re, G. D.; Minichino, C. *J. Theor. Biol.* 1985, *115*, 571.
(69) Mullay, J. *J. Am. Chem. Soc.* 1986, *108*, 1770.
(70) Chirlian, L. E.; Francl, M. M. *J. Comp. Chem.* 1987, *8*, 894.
(71) Richards, W. G. *Quantum Pharmacology*, 2nd ed.; Butterworth & Co.: London, 1983.
(72) Cohen, N. C.; Colin, P.; Lemoine, G. *Tetrahedron* 1981, *37*, 1711.
(73) Smith, G. M.; Veber, D. F. *Biochem. Biophys. Res. Commun.* 1986, *134*, 907.
(74) Dolata, D. P.; Leach, A. R.; Prout, K. *J. Comput.-Aided Mol. Des.* 1987, *1*, 73.
(75) Lipton, M.; Still, W. C. *J. Comp. Chem.* 1988, *9*, 343.

Distance geometry produces a random sampling of conformation space by selecting random distances within each pair of upper and lower bounds. This approach samples conformation space rapidly and efficiently, but cannot guarantee that all of conformation space has been searched. Systematic dihedral search methods can in theory promise that all conformation space is adequately searched, but in practice, the completeness of the search is limited by the increment used in the dihedral scan. The time required for systematic search increases exponentially with each additional rotatable bond and becomes impractical beyond 12–13 rotatable bonds. The time required for distance geometry is independent of the number of rotatable bonds and depends only on the total number of atoms; distance geometry has approximately a quadratic time dependence on the number of atoms and therefore is still practical for large structures that are beyond the reach of systematic search methods. Cyclic structures are handled naturally by distance geometry with no decrease in efficiency, but systematic search method must deal with the ring-closure problem which further limits their efficiency and range.[73] Both methods require molecular mechanics calculations to calculate the energy of each generated conformation; systematic search methods often use a single-point energy calculation since bond lengths and angles are not distorted from their ideal values, but distance geometry requires at least partial energy minimization since all degrees of freedom are varied. Distance geometry is currently not available in any major molecular modeling software system, but stand-alone programs are available commercially,[76] from QCPE[53,77] or from UCSF.[78]

The ellipsoid algorithm is a promising new approach for generating low-energy conformations of molecules by efficiently sampling among the sterically allowed combinations of dihedral angles. It has been applied to the conformational analysis of 18-crown-6,[79] the determination of peptide solution structure using NMR distance constraints,[47] and ligand–protein docking.[48] For small to medium-sized molecules it may be more efficient than either systematic search or distance geometry for locating deep energy minima.

(f) Physical Properties. Although conformational analysis constitutes one important aspect of molecular modeling, a number of physical properties are also accessible with theoretical calculations. Molecular mechanics, semiempirical, and ab initio methods[42] can give rather reliable results on various molecular properties such as heats of formation, enthalpies (e.g. in evaluating the relative stability of isomers), barriers and activation energies, dipole moments, reaction paths, etc. Theoretical calculations can provide a number of indices that may not be directly related to experimental data but that can be very useful because they carry high physical information content (molecular, localized, and frontier orbitals, electronegativities, polarization, delocalization, atomic and bond population, etc.). For example, electron densities are useful because they provide a good basis for the analysis of the stereoelectronic properties of either isolated or interacting molecules. Molecular electrostatic potentials are usually generated from the partial atomic charges derived from a quantum mechanical calculation. Most of the major software systems include facilities to calculate and display electrostatic potentials. Other properties can be calculated by empirical methods; the most popular are the prediction of log P (octanol/water partition coefficient) and MR (molar refractivity) as developed by the Pomona College Medicinal Project.[80,81]

IV. Modeling Sets of Small Molecules

In indirect drug design the modeling is based on the recognition of three-dimensional stereochemical features common to sets of active molecules—the pharmacophore. Superposition and comparison methods, often called "molecular fitting" or "pharmacophore alignment", are the most routinely available. They compare, on a pairwise basis, an active reference compound with a set of other structures. Excluded volume analysis[82] is a classical way to geometrically compare a set of active and inactive molecules in order to reveal essential features, based on the simple idea that regions of inactive molecules which protrude beyond the volume common to the active molecules indicate sterically unfavorable regions on the receptor. The most popular approach to phamacophore superimposition has been the "active analogue" approach, developed by Marshall et al.[83,84] which uses systematic search to determine the allowed conformations of all molecules in the study, followed by comparison of interatomic distances to select conformers that overlap, based on the proposed pharmacophore. Attempts to take into consideration the conformational energies during the fitting process have been made.[85,86] The more recent "ensemble distance geometry method[77,87] will rapidly determine if any solutions exist without replacing a complete systematic search and, if so, provide a random sampling of solutions that indicates how uniquely determined the model is. Additional advantages of this approach are that it handles rings naturally without the ring closure difficulties encountered in dihedral search methods and that chirality can be allowed to vary for any stereo centers of unknown absolute configuration.

Most available systems provide simple interactive fitting functionality by considering the molecules as conformationally rigid, while optionally allowing motion of a few dihedral angles.[85,88] Most of the major software systems

(76) Hare, D. *DSPACE*, Infinity Systems: 14810 216th Ave. NE, Woodinville, WA 98072, 1988.
(77) Blaney, J. M.; Crippen, G. M. *DGEOM, to be submitted*, Quantum Chemistry Program Exchange, Indiana University: Bloomington, 1990.
(78) Kuntz, I. D.; Crippen, G. M. *EMBED*, Department of Pharmaceutical Chemistry, University of California, San Francisco: San Francisco, CA 94143, 1980.
(79) Billeter, M.; Howard, A. E.; Kuntz, I. D.; Kollman, P. A. *J. Am. Chem. Soc.* **1988**, *110*, 8385.
(80) Hansch, C.; Leo, A. *Substituent Constants for Correlation Analysis in Chemistry and Biology*; John Wiley and Sons: New York, 1979.
(81) Leo, A.; Weininger, D.; Weininger, A. *CLOGP, CMR*, Medicinal Chemistry Project, Pomona College: Claremont, CA 91711; version 3.54, distributed by Daylight Chemical Information Systems, 1989.
(82) Sufrin, J. R.; Dunn, D. A.; Marshall, G. R. *Mol. Pharmacol.* **1981**, *19*, 307.
(83) Marshall, G. R.; Barry, C. D.; Bosshard, H. E.; Dammkoehler, R. A.; Dunn, D. A. In *Computer-Assisted Drug Design*; Olson, E. C., Christofferson, R. E., Eds.; ACS Symposium Series 112; American Chemical Society: Washington, DC, 1979; p 205.
(84) Marshall, G. R.; Motoc, I. In *Molecular Graphics and Drug Design, Topics in Molecular Pharmacology*; Burgen, A. S. V., Roberts, G. C. K., Tute, M. S., Eds.; Elsevier: Amsterdam, 1986; Vol. 3, p 115.
(85) Labanowski, J.; Motoc, I.; Naylor, C. B.; Mayer, D.; Dammkoehler, R. A. *Quant. Struct.-Act. Relat.* **1986**, *5*, 138.
(86) Cohen, N. C. In *Computer-Assisted Drug Design*: Olson, E. C., Christoffersen, R. E., Eds.; ACS Symposium Series 112; American Chemical Society: Washington, DC, 1979; p 371.
(87) Sheridan, R. P.; Nilakantan, R.; Dixon, J. S.; Venkataraghavan, R. *J. Med. Chem.* **1986**, *29*, 899.
(88) Cory, M.; Bentley, J. *J. Mol. Graphics* **1984**, *2*, 39.

have integrated flexible fit computational modules in which not only the internal rotational degrees of freedom but also the conformational energies of the individual molecules are taken into account. MAXIMIN[85] is an example in which two alternative methods are possible: a set of flexible molecules can be mapped onto a rigid reference compound, or all the molecules are treated as flexible entities, and the treatment is directed toward the minimization of the conformational variance of the whole set. "Template forcing"[89] is another way to maximize overlaps between molecules using restrained molecular mechanics and dynamics.

In molecular fitting treatments the maximization of the overlaps is generally achieved by geometrical least-squares minimizations, which requires a preliminary selection of pairs of atoms expected to be superimposable. The choice of the pairs of atoms is very subjective, on the basis of "chamber intuition" and the hypothesized pharmacophore. Less subjective approaches have also been developed, on the basis of maximizing the overlap of a set of molecules by minimizing the exposed area of the entire set while simultaneously ensuring that the energies of the individual molecules remain close to a local minimum,[86] combinatorial methods for comparing all possible overlaps of similar atom types,[90,91] and approaches based on three-dimensional electrostatic potential similarity,[92,93] molecular surface similarity,[94] and molecular shape analyses.[161-164]

A more physical approach is to force common pharmacophore atoms to interact with a common binding site, defined by hypothetical points of interaction (e.g. dummy atoms), rather than forcing them to directly superimpose. Different chemical moieties can be compared and do not need to be exactly superimposable.[155,156] Several systems provide Boolean logical operators (and, or, not, etc.) which allow one to find common similarities between two molecules in terms of preselected electrostatic contours or molecular volumes. Cramer et al.[95] recently described a promising new 3D-QSAR method based on calculating the interaction of each molecule in a set of superimposed active structures with a variety of probe atoms on a three-dimensional lattice.

New approaches developed on databases of minimized conformers and using three-dimensional substructure and similarity search techniques[30] have already shown value in identifying pharmacophoric moieties and associated active conformations of molecules.[33] Efforts of this type are current topics of modeling development and are just now becoming available.

V. Macromolecule Modeling

X-ray crystallography and macromolecular modeling provide the most detailed possible view of drug–receptor interactions and have created a new, rational approach to drug design where the structure of a drug is designed on the basis of its fit to the three-dimensional structure in the receptor site, rather than by analogy to other active structures or random leads.[96,97] There are now over 300 X-ray crystal structures of proteins and nucleic acids that have been solved; most are available in the Brookhaven Protein Data bank,[98] including several ligand–macromolecule complexes. Although relatively few structures of actual or potential drug receptors have been solved, the rate of solving these structures has increased steadily during the last few years and will continue to increase due to improvements in crystallographic techniques and the availability of new protein through recombinant DNA approaches. Such high-resolution structures offer the potential of designing drugs tailor-made to fit their receptor with high affinity and selectivity. However, the rate of release to the public domain of three-dimensional coordinates of important macromolecules is decreasing even as the rate of solving them increases. The results of the technology that promised this great potential for rational, receptor-based drug design are in fact often not available. The issues surrounding this counterproductive situation have been discussed previously.[99,100]

Despite the impressive advances in macromolecular X-ray crystallography, availability of high-quality crystals remains the major limiting factor. 2D NMR techniques have advanced tremendously[56,101,102] and can now provide three-dimensional structural information on small proteins (up to 100–150 residues) and DNA in solution, using distance geometry[53,54] and/or restrained molecular dynamics[52,103] to build models consistent with distance constraints derived from NOE (nuclear overhauser enhancement) and coupling constant data.[55] In several cases 2D NMR has been used to solve a complete protein structure; Tendamistat, the 75-residue α-amylase inhibitor, was solved independently by 2D NMR[104,105] and X-ray crystallography,[106] resulting in very similar structures. 2D NMR previously provided only low-resolution models that revealed the overall folding pattern with little information about side-chain locations, but Wuthrich's group has recently determined the complete solution structure of Tendamistat by NMR, including all side chains[103]. The January 1989 release of the Brookhaven Protein Data Bank[98] includes for the first time a protein structure solved in solution by NMR; other structures solved by NMR will follow.

Most current software systems provide efficient means for the construction of polymeric fragments. Peptides, nucleic acids, or carbohydrates are easily generated in an arbitrary or user-defined three-dimensional conformation by selecting in a menu the linear sequence combined with additional information indicating how the progressively growing molecule should fold. The growth either can be fully extended or can follow commonly observed secondary structure (e.g. α-helix, β-sheet in the case of peptides; A,

(89) Struthers, R. S.; Rivier, J.; Hagler, A. T. In *Conformationally based drug design: Peptides and nucleic acids as templates or targets*; Vida, J. A., Gordon, M., Ed.; American Chemical Society: Washington, DC, 1984; p 239.
(90) Crippen, G. M. *J. Med. Chem.* **1980**, *23*, 599.
(91) Danziger, D. J.; Dean, P. M. *J. Theor. Biol.* **1985**, *116*, 215.
(92) Dean, P. M.; Chau, P. L. *J. Mol. Graphics* **1987**, *5*, 152.
(93) Namasivayam, S.; Dean, P. M. *J. Mol. Graphics* **1986**, *4*, 46.
(94) Dean, P. M.; Callow, P.; Chau, P. L. *J. Mol. Graphics* **1988**, *6*, 28.
(95) Cramer, R. D., III; Patterson, D. E.; Bunce, J. D. *J. Am. Chem. Soc.* **1988**, *110*, 5959.
(96) Beddell, C. R. *Chem. Soc. Rev.* **1984**, *13*, 279.
(97) Hol, W. G. J. *Angew. Chem.* **1986**, *25*, 767.
(98) Bernstein, F. C.; Koetzle, T. F.; Williams, G. T. B.; Meyer, E. F.; Brice, M. D.; Brodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. *J. Mol. Biol.* **1977**, *112*, 535.
(99) Richards, F. M. *J. Comput.-Aided Mol. Des.* **1988**, *2*, 3.
(100) Marshall, G. R.; Vinter, J. G.; Holtje, H. D. *J. Comput.-Aided. Mol. Des.* **1988**, *2*, 1.
(101) Kaptein, R.; Boelens, R.; Scheek, R. M.; van Gunsteren, W. F. *Biochemistry* **1988**, *27*, 5389.
(102) Wuthrich, K. *Science* **1989**, *243*, 45.
(103) de Vlieg, J.; Scheek, R. M.; van Gunsteren, W. F.; Berendsen, H. J. C.; Kaptein, R.; Thomason, J. *Proteins* **1988**, *3*, 209.
(104) Kline, A. D.; Braun, W.; Wuthrich, K. *J. Mol. Biol.* **1986**, *189*, 377.
(105) Kline, A. D.; Braun, W.; Wuthrich, K. *J. Mol. Biol.* **1988**, *204*, 675.
(106) Pflugrath, J. W.; Weigand, G.; Huber, R. *J. Mol. Biol.* **1986**, *189*, 383.

B, or Z conformation for nucleic acids, and analogous prespecified conformers for carbohydrates). These simple methods have little chance of leading to meaningful three-dimensional structures unless they are used in combination with additional knowledge and experimental data.

Many more protein sequences are available than crystal structures, and the gap will continue to grow as DNA-sequencing methods become even faster. Fortunately, protein sequences occasionally show high sequence homology with proteins whose three-dimensional structure is known, suggesting the possibility of modeling the unknown structure based on the crystal structure of the homologous protein. This has become a popular approach and has recently been reviewed by Blundell et al.;[107,108] an example is the recent prediction of the three-dimensional structure of tissue plasminogen activator.[109] Homology modelling techniques have been particularly successful for predicting antibody structures.[110,111] Jones and Thirup[112] showed that it may be possible to fit most secondary structure elements using fragments from other proteins of known structure; this approach is useful for building models for insertion and deletion regions and for homology model building in general. Most of the macromolecular modeling software systems contain similar facilities for protein homology modeling.

For the majority of protein sequences with little significant homology to known structures, the problem of predicting secondary and tertiary structure accurately enough for drug design applications is still insurmountable.[113] Error rates for the various secondary structure prediction approaches are usually greater than 40%.[114,115] However, several of the current methods can suggest families of possible secondary structures that may be useful for some applications (e.g. site-directed mutagenesis). Few predictions of complete secondary and tertiary structure have been reported. A realistic appraisal of the current state of the art is represented by Cohen et al.'s ambitious prediction[116] of the core tertiary structure of Interleukin-2 prior to its determination by X-ray crystallography;[117] while the prediction had several key features correct, it was too inaccurate to be useful for drug design[118]—even small errors in the placement of secondary and tertiary structure can lead to major errors in the complete model.

VI. Modeling Drug–Receptor Interactions

The major interactions involved in drug–receptor binding are electrostatic (including hydrogen bonding), dispersion or van der Waals, and hydrophobic.[119] Hydrophobic interactions usually provide the major driving force for binding, while hydrogen-bonding and electrostatic interactions primarily provide specificity and often add little to the free energy f binding.[120-122] Drug–receptor "docking" is typically done interactively with molecular surface displays (e.g. "extra radius" surface) used to guide the fit, based on hydrophobic or electrostatic potential color coding. Since it is difficult to hit a m ving target, the binding site is usually treated as completely rigid initially, while the conformation of the ligand is adjusted interactively. Recent systems are fast enough to provide real-time energy calculations while docking (future systems may use this information to provide feedback and prevent steric collisions or high-energy conformations). High-energy contacts can be shown with color-coded vectors.[123] Interactive docking thus alternates between continuous motion, possibly with real-time updates of the interaction energy if fast hardware is available, and periodic cycles of energy minimization to clean up the visual fit. A simple feedback approach that scales the dial (or joystick) response based on the instantaneous derivative of the interaction energy facilitates docking.[124] If the user moves uphill in energy, the system resists the motion, but if the user is moving in a favorable direction, the system encourages the motion by increasing responsiveness, so the docking tends to follow the path of least resistance in a sort of interactive energy minimization. Finally, energy minimization of the entire complex, where all atoms are allowed to relax, provides a good indication of the plausibility of the model and a rough estimate of the relative interaction enthalpy of the candidate drug. Ionic interactions and hydrogen bond energies are usually overestimated in a typical calculation due to the omission of solvent hydrogen-bonding competition; these effects are treated properly in the free energy perturbation the ry method described below.

Conventional energy minimization with this many degrees of freedom is easily trapped in local minima and can give deceptive results; energy minimization rarely produces a structure that is significantly different from the starting coordinates. Molecular dynamics simulations as short as 10 ps are much better at escaping local minima and can give much lower energy structures; a good strategy is to begin with a short dynamics run and follow it with energy minimization. Such short dynamics simulations contain no meaningful information about the actual motions or dynamics of the structure (up to 30 ps may be required just for thermal equilibration); they simply provide a more efficient method of energy minimization and a good indication of the stability of the model (poor models tend to fly apart very quickly).

Multiple binding modes are often possible, as shown by the X-ray structure of an elastase-product complex in which the ligand is bound backwards to the established mode of productive binding.[125] It can be very difficult with interactive methods to find the most likely binding

(107) Blundell, T. L.; Sibanda, B. L.; Sternberg, M. J. E.; Thornton, J. M. *Nature* 1987, *326*, 347.
(108) Blundell, T.; Carney, D.; Gardner, S.; Hayes, F.; Howlin, B.; Hubbard, T.; Overington, J.; Singh, D. A.; Sibanda, B. L.; Stuckliffe, M. *Eur. J. Biochem.* 1988, *172*, 513.
(109) Heckel, A.; Hasselbach, K. M. *J. Comput.-Aided. Mol. Des.* 1988, *2*, 7.
(110) Chothia, C.; Lesk, A. M. *J. Mol. Biol.* 1987, *196*, 901.
(111) Bruccoleri, R. E.; Haber, E.; Novotny, J. *Nature* 1988, *335*, 564.
(112) Jones, T. A.; Thirup, S. *EMBO J.* 1986, *5*, 819.
(113) Yada, R. Y.; Jackman, R. L.; Nakai, S. *Int. J. Peptide Protein Res.* 1988, *31*, 98.
(114) Kabsch, W.; Sander, C. *FEBS Lett.* 1983, *155*, 179.
(115) Nishikawa, K. *Biochim. Biophys. Acta* 1983, *748*, 285.
(116) Cohen, F. E.; Kosen, P. A.; Kuntz, I. D.; Epstein, L. B.; Ciardelli, T. L.; Smith, K. A. *Science* 1986, *234*, 349.
(117) Bandhuber, B. J.; Boone, T.; Kenney, W. C.; McKay, D. B. *Science* 1987, *238*, 1707.
(118) Landgraft, B.; Cohen, F. E.; Smith, K. A.; Gadski, R.; Ciardelli, T. L. *J. Biol. Chem.* 1989, *264*, 816.
(119) Kollman, P. A. In *X-ray Crystallography and Drug Action*; Horn, A. S., Ranter, C. J. D., Eds.; Oxford University Press: Oxford, 1984; p 63.
(120) Fersht, A. R. *TIBS* 1984, *9*, 145.
(121) Fersht, A. R.; Shi, J.; Knill-Jones, J.; Lowe, D. M.; Wilkinson, A. J.; Blow, D. M.; Brick, P.; Carter, P.; Waye, M. M. Y.; Winter, G. *Nature* 1985, *314*, 235.
(122) Street, I. P.; Armstrong, C. R.; Withers, S. G. *Biochemistry* 1986, *25*, 6021.
(123) Bush, B. L. *Comput. Chem.* 1984, *8*, 1.
(124) Swanson, E.; Blaney, J. M. Unpublished results.
(125) Meyer, E. F., Jr.; Radhakrishnan, R.; Cole, G. M.; Presta, L. G. *J. Mol. Biol.* 1986, *189*, 533.

mode candidates. Naruto et al.[126] used a systematic search procedure to find chymotrypsin tetrahedral intermediate conformers given a covalent bond linking the ligand with the site. DesJarlais et al.[127] developed a general docking method for conformationally flexible ligands based on a fast sphere-matching algorithm by docking each rigid fragment of the ligand (fragments between rotatable bonds) independently.

A major problem with all design approaches is our current lack of ability to calculate even a qualitatively accurate estimate of the free energy of binding between two molecules in aqueous solution. An important advance in modeling ligand–receptor interactions is the recent application of free energy perturbation methods.[129,130] This takes advantage of the properties of a thermodynamic cycle to simulate a physical process which is very difficult to calculate (the transfer of a drug from solution into a receptor binding site, compared with the transfer of its analogue) by an equivalent nonphysical process (the "mutation" of a drug into its analogue, performed both in solution and in the binding site) which is relatively easy to calculate. This "mutation" is carried out by gradually changing the parameters of the initial drug molecule to the parameters of the final drug molecule during a molecular dynamics simulation, which is performed once in "solution", usually in a box of several hundred water molecules, and again in the macromolecule. The simulation starts with 100% initial drug character and ends with 100% final drug character; intermediate steps in the simulation have nonphysical hybrid drug molecules. Molecular dynamics generates a statistical mechanical ensemble average at each point along the simulation as the properties of the initial molecule are varied. Such simulations require large amounts of supercomputer time.

Wong and McCammon[131] described the calculation of the free energy difference of binding benzamidine vs *p*-fluorobenzamide to trypsin, while Bash et al.[132] reported calculations on free energy of binding differences for several thermolysin inhibitors and for a single thermolysin inhibitor to different mutant thermolysins. Both simulations were accurate to within less than 1 kcal of the experimental value. These results demonstrated how important the role of differential solvation can be in determining binding-affinity differences. It is not clear yet how large a difference between molecules can be simulated; all drug–receptor simulations so far have involved conservative single atom replacements, although Singh et al.[133] found excellent results with changes in entire amino acid side chains for calculating differences in solvation free energy. Free-energy perturbation methods are gradually becoming available in several molecular modeling systems, although this is still a frontier research area and it is not clear what the best approaches are or how long a simulation must be run to ensure statistically significant results.

Free energy perturbation methods offer the exciting possibility of calculating accurate differences in binding free energies between related ligands, which could make it possible to predict the binding affinity of new compounds prior to synthesis. Merz and Kollman[188] recently demonstrated the predictive ability of the approach by estimating the $\Delta(\Delta G)$ of thermolysin binding to a new inhibitor. However, recent work[189,190] has pointed out that it is extremely difficult to verify when a simulation has converged and has shown that some of the early reports were rather optimistic and tended to overestimate the precision with which $\Delta(\Delta G)$ was calculated. It is now clear that additional basic research is necessary before the method can be routinely applied and yield quantitatively reliable results. Current results suggest that $\Delta(\Delta G)$ for ligand–macromolecule binding can be calculated to within ±1.5–2 kcal/mol (equivalent to about a factor of 10–30 in binding affinity). Van Gunsteren[189], and Pearlman and Kollman[190] reviewed problems and pitfalls of the approach recently.

VII. Design

In the past, drugs were designed with an almost total naivete from the point of view of the molecular mechanisms of the underlying molecular machinery involved. The recent developments in Molecular Biology have clearly revealed the critical importance of three-dimensionality (3D) in molecular recognition and discrimination aspects. Even when the 3D features of the biological proteins involved were not known, drug design conducted along with this line emerged as an important aim and stimulated the development of some of the techniques mentioned in paragraph IV. Examples of lead molecules conceived in this way have been regularly reviewed,[1,5,184] and it is beyond the scope of this article to review all the excellent contributions that were made in this perspective.

As far as direct drug design is concerned, the ability to model both small organic molecules and macromolecules in the same system is critical; several of the systems currently available were originally designed for handling the regular, repeating polymeric structure of proteins and nucleic acids and deal rather poorly with the more arbitrary structures found in small organic molecules. Others were initially designed for modeling small molecules and do not handle macromolecular structures well. Few systems come close to offering the best of macromolecular and small-molecule modeling in an integrated system, providing the ability to interactively design and build potential ligands directly into a macromolecular receptor binding site.

Computer graphics enables us to qualitatively visualize drug–receptor interactions and molecular mechanics can provide rough estimates of the interaction energy, which allow us to design molecules that are apparently complementary to a binding site. For close analogues this can be sufficient to both rationalize the relative activities of a series of analogues and design new, closely related analogues; several excellent examples of this approach have been reported.[96,97,134] An integrated approach[135] combining molecular modeling with QSAR has proven to be especially powerful for this application, since the QSAR can help differentiate between different possible binding modes. We have much less experience in the de novo design of novel molecules (without a lead compound in an X-ray structure with its receptor). The designs by Beddell et al. of 2,3-diphosphoglycerate mimics[136] and antisickling com-

(126) Naruto, S.; Motoc, I.; Marshall, G. R.; Daniels, S. B.; Sofia, M. J.; Katzenellenbogen, J. A. *J. Am. Chem. Soc.* 1985, *107*, 5262.
(127) DesJarlais, R. L.; Sheridan, R. P.; Dixon, J. S.; Kuntz, I. D.; Venkataraghavan, R. *J. Med. Chem.* 1986, *29*, 2149.
(128) Chang, G.; Still, W. C.; Guida, W. C. *J. Am. Chem. Soc.* 1989, *111*, 4379.
(129) van Gunsteren, W. F.; Berendsen, H. J. C. *J. Comput.-Aided Mol. Des.* 1987, *1*, 171.
(130) McCammon, J. A. *Science* 1987, *238*, 486.
(131) Wong, C. F.; McCammon, J. A. *J. Am. Chem. Soc.* 1986, *108*, 3830.
(132) Bash, P. A.; Singh, U. C.; Brown, F. K.; Langridge, R.; Kollman, P. A. *Science* 1987, *235*, 574.
(133) Singh, U. C.; Brown, F. K.; Bash, P. A.; Kollman, P. A. *J. Am. Chem. Soc.* 1987, *109*, 1607.
(134) Roth, B. *Fed. Proc., Fed. Am. Soc. Exp. Biol.* 1986, *45*, 2765.
(135) Hansch, C.; Klein, T. *Acc. Chem. Res.* 1986, *19*, 392.

pounds[137] based on the hemoglobin X-ray structure are still some of the best examples of this approach, despite the fact that most of this work was done with wire models! The only other reported successful example of de novo design using computer modeling methods is the design of phospholipase A_2 inhibitors by Ripka et al.[138]

All of the approaches we have described so far are analytical and oriented toward modeling known structures. Where do the structures of novel candidate drugs come from? Actual molecular structure design is still a formidable challenge dependent on the creativity, ingenuity, and experience of the medicinal chemist. Goodford developed a simple molecular mechanics based approach for calculating optimal ligand atom locations in a binding site, which is an important first step.[139] The method is based on calculating the interaction energy for each of a variety of probes (e.g. hydroxyl oxygen, carbonyl oxygen, carboxyl oxygen, amide nitrogen, amine nitrogen, etc.) at each point on a three-dimensional grid superimposed on the binding site. The grid is then contoured by energy, and the resulting contours are graphically displayed (as color-coded contour maps or dot clouds) in the binding site. The contours indicate predicted "hot spots" where a ligand atom of a given type should prefer to bind. Unfortunately it is usually very difficult to connect each of these "hot spots" together into a synthetically accessible molecule in a low-energy conformation, but the method does provide useful visual clues for structure design.

Current design techniques combine Goodford's (or related methods) with the other previously described interactive methods, where the investigator fits a variety of organic fragments in a trial and error fashion into the site, attempting to eventually combine the fragments into a complete molecule. The best approach is usually to design and build the developing ligand piece by piece in the binding site by combining preformed fragments from a library of different ring systems and functional groups and/or with CONCORD.[28] Small molecules can be built rapidly this way, and the resulting structures are usually accurate enough for initial qualitative "docking" into the site model. This is where good interactive software design and a well-thought-out user interface are especially important, since the modeler will spend much of his time in this stage trying out new ideas. Although it seems likely that all the information required for the design of an optimal ligand is present in the high-resolution structure of the receptor site, no systematic approaches exist yet for complete de novo design. The sphere-matching flexible ligand docking approach of DesJarlais et al.[127] or a 3D pharmacophore search over a 3D database[20,33,140] may eventually be able to achieve this, by docking fragments from a large library and then combining the fragments into complete molecules.

Very recently Dean and Colleagues[165-168] have published exploratory investigations concerning the possibility of automated site-directed drug design. The aim is to conceive appropriate algorithms and to construct a knowledge base for the automatic construction of novel ligands to fit specified binding sites.

(136) Beddell, C. R.; Goodford, P. J.; Norrington, F. E.; Wilkinson, S.; Wootton, R. *Br. J. Pharmacol.* **1976**, *57*, 201.
(137) Beddell, C. R.; Goodford, P. J.; Kneen, G.; White, R. D.; Wilkinson, S.; Wootton, R. *Br. J. Pharmacol.* **1984**, *82*, 397.
(138) Ripka, W. C.; Sipio, W. J.; Blaney, J. M. *Lect. Heterocycl. Chem.* **1987**, *IX*, S95.
(139) Goodford, P. J. *J. Med. Chem.* **1985**, *28*, 849.
(140) Van Drie, J. H.; Weininger, D.; Martin, Y. C. *J. Comp.-Aided. Mol. Des.*, submitted.

VIII. Molecular Modeling Software

Major currently available academic and commercial molecular modeling software systems are listed below, along with their major functions. Currently, available computer (PC) programs have limited functionality for medicinal chemistry applications; they have not been included in this paper. Gerson[175] and Sadek[187] recently compiled reviews of PC software available for basic molecular modeling applications.

Tripos[141] has developed an excellent PC (IBM PC or Apple Macintosh II) interface to the host software (running on a superminicomputer or workstation) using the PC's local processing to provide real-time graphics display and manipulation of up to 100–200 atoms. This approach, which is now appearing in an increasing number of modeling packages, takes advantage of the inexpensive, fast graphics performance of the latest generation of PC's for display of small to medium-sized molecules, but retains the full functionality of the host software on a larger computer.

Program	Functions[a]
AMBER[36]	M, MM, MD, FE
BIOGRAF[142]	G, S, M, CA, MM, MD, MO
CHEM-X[31]	G, S, M, CA, MM, STAT, MO
CONCORD[28]	S
DISGEO[53]	DG
DISMAN[54]	DG
DSPACE[70]	DG
EMBED[78]	DG
FRODO[143,144]	G, M
GRID[139,145]	PR
GROMOS[146]	M, MM, MD, FE
INSIGHT/DISCOVER/DELPHI[43]	G, S, M, CA, MM, MD, MO
MACROMODEL[147,148,157]	G, S, M, CA, MM, MD, MO
MIDAS[149,150]	G, M
MM2[24]	MM, CA
MOGLI[151]	G, S, M
QUANTA/CHARMM[29,152]	G, S, M, CA, MM, MD, FE, PR, STAT, MO
SYBYL/ALCHEMY/NITRO[141]	G, S, M, CA, MM, MD, STAT, MO

[a] G graphic display and manipulation
S Small molecule structure building
M Macromolecules structure building
CA Conformational analysis facilities
MM Molecular mechanics
MD Molecular dynamics
FE Free energy perturbation methods
DG Distance geometry
PR Probe interaction energies
STAT Statistical tools
MO Molecular orbital methods from QCPE

(141) Tripos Associates, St. Louis, MO 63117.
(142) BioDesign, 199 South Los Robles Ave., Pasadena, CA 91101.
(143) Jones, T. A. *J. Appl. Crystallogr.* **1978**, *11*, 268.
(144) Jones, T. A. In *Computational Crystallography*; Sayre, D., Ed.; Clarendon Press: Oxford, 1982; p 303.
(145) Goodford, P. J. *GRID*, Molecular Discovery Ltd., West Way House, Elms Parade: Oxford OX2 9LL, England, 1986.
(146) van Gunsteren, W. F.; Berendsen, H. J. C. *Groningen Molecular Simulation (GROMOS)*, Biomos: Nijenborgh 16, 9747 AG Groningen, The Netherlands, 1987.
(147) Still, W. C.; MacPherson, L. J.; Harada, T.; Callahan, J. F.; Rheingold, A. L. *Tetrahedron* **1984**, *40*, 2775.
(148) Still, W. C. *Macromodel*, Department of Chemistry, Columbia University: New York, 1984.
(149) Ferrin, T. E.; Huang, C. C.; Jarvis, L. E.; Langridge, R. *J. Mol. Graphics* **1988**, *6*, 2.
(150) Ferrin, T. E.; Huang, C. C.; Jarvis, L. E.; Langridge, R. *J. Mol. Graphics* **1988**, *6*, 13.
(151) MOGLI, Evans & Sutherland Computer Corporation, Salt Lake City, UT.

IX. Perspective

Crystallographers pioneered techniques to visualize, scrutinize, and manipulate three-dimensional molecular models. For example, the ORTEP[153] program plots crystal structure illustrations. ORTEP is still widely valued, in particular to add the third-dimension perspective to molecular structure representations. Another early example of a macromolecular graphics system is FRODO,[143,144] a software program used to facilitate electron density fitting experiments and to display and examine protein structures.

Quite independently, early attempts to incorporate computational chemistry methods to study the properties of molecules of biological interest have appeared in software such as, for example, AMBER,[36] CHARMM,[39] PCILO,[63,64] MM2,[34] and CAMSEQ.[154]

It was not until later, however, that molecular modeling graphics systems emerged from the combination of the above techniques and methods. With the addition of a conformational dimension to support structure–activity studies, the medicinal chemist was progressively offered an expanding arsenal of tools to assist and enhance drug design attempts. As outlined in this review, there is now an ample choice of molecular modeling software and methods available to the medicinal chemist.

Initial modeling software packages have been designed to provide methods dedicated either to small organic molecule or macromolecular modeling applications. Recently, progress has been made in combining both applications in a single package. However, a better integration of these two aspects is still needed to improve compatibility and enhance user interaction. In addition, future developments should benefit from a concerted combination of strengths in specific techniques and methodologies, particularly when addressing the increasing number of applications for the study of the interactions between small organic molecules and macromolecules.

Recent evolution in hardware and software technologies has made possible both implementation and development of methods (e.g., molecular dynamics, real-time manipulation of colored solid-shaded images for macromolecules) that were prohibitive not so long ago. Simultaneously, software packages have progressed to take advantage of powerful state-of-the-art features (e.g., windowing, menu-driven systems, command language syntax). However, the desirable user-friendly interface has been somewhat overlooked in this evolutionary process, and modeling software can appear rather complex and cumbersome to occasional users. We hope that future developments will address this issue.

Advances in molecular modeling have been impressive over the last years. Major milestones in software and hardware technologies have been accomplished and future prospects in this rapidly evolving arena look very promising. Current efforts to develop and integrate methods and techniques to assist and enhance drug design studies should lead to even higher levels of computer automation, rationalization, quantification, and, eventually, de novo design of novel molecules.

(152) Polygen Corporation, 200 Fifth Ave., Waltham, MA 02254.
(153) Johnson, C. K. *ORTEP-II: A Fortran Thermal-Ellipsoid Plot Program for Crystal Structure Illustration*; Oak Ridge National Laboratory, ORNL-3794, UC-4-chemistry; Oak Ridge, TN, June 1965.
(154) Potenzone, R., Jr.; Cavicchi, E.; Weintraub, H. J. R.; Hopfinger, A. J. *Comput. Chem.* **1977**, *1*, 187.
(155) Kato, Y.; Itai, A., Iitaka, Y. *Tetrahedron* **1987**, *22*, 5229.
(156) Itai, A.; Kato, Y.; Tomioka, N.; Iitaka, Y.; Endo, Y.; Hasegawa, M.; Shudo, K.; Fujiki, H.; Sakai, S. I. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 3688.
(157) Mohamadi, F.; Richards, N. G. J.; Guida, W. C.; Liskamp, R.; Lipton, M.; Caufield, C.; Chang, G.; Hendrickson, T.; Still, W. C. *J. Comput. Chem.* **1989**, submitted.
(158) Gerber, P. R.; Gubernator, K.; Müller, K. *Helv. Chim. Acta* **1988**, *71*, 1429.
(159) Hoflack, J.; De Clercq, P. J. *Tetrahedron* **1988**, *44*, 6667.
(160) Gund, P.; Barry, D. C.; Blaney, J. M.; Cohen, N. C. *J. Med. Chem.* **1988**, *31*, 2230.
(161) Hopfinger, A. J. *J. Am. Chem. Soc.* **1980**, *102*, 7196.
(162) Hopfinger, A. J. *J. Med. Chem.* **1981**, *24*, 818.
(163) Hopfinger, A. J. *J. Med. Chem.* **1983**, *26*, 990.
(164) Hopfinger, A. J. *J. Med. Chem.* **1985**, *28*, 946.
(165) Danziger, D. J.; Dean, P. M. *Proc. R. Soc. London* **1989**, *236*, 101.
(166) Danziger, D. J.; Dean, P. M. *Proc. R. Soc. London* **1989**, *236*, 115.
(167) Lewis, R. A.; Dean, P. M. *Proc. R. Soc. London* **1989**, *236*, 125.
(168) Lewis, R. A.; Dean, P. M. *Proc. R. Soc. London* **1989**, *236*, 141.
(169) Scheraga, H. A. *Prog. Clin. Biol. Res.* **1989**, *289*, 3.
(170) Gibson, K. D.; Scheraga, H. A. *J. Comput. Chem.* **1987**, *8*, 826.
(171) Purisima, E. O.; Scheraga, H. A. Proc. *Natl. Acad. Sci. U.S.A.* **1986**, *83*, 2782.
(172) Li, Z.; Scheraga, H. A. Proc. *Natl. Acad. Sci. U.S.A.* **1987**, *84*, 6611.
(173) Momany, F. A.; McGuire, R. F.; Burgess, A. W.; Scheraga, H. A. *J. Phys. Chem.* **1975**, *79*, 2361.
(174) Nemethy, G.; Pottle, M. S.; Scheraga, H. A. *J. Phys. Chem.* **1983**, *87*, 1883.
(175) Gerson, C. K. *Chemical Structure Software for Personal Computers*; Meyer, D. E., Warr, W. A., Love, R. A., Eds.; American Chemical Society: Washington, DC, 1988; p 53.
(176) MacKay, D. H. J.; Cross, A. J.; Hagler, A. T. *Prediction of Protein Structure and The Principle of Protein Conformation*; Fasman, G., Ed.; Plenum Press: New York, 1989; p 317.
(177) Burt, S. K.; MacKay, D.; Hagler, A. T. *Computer-Aided Drug Design, Methods and Applications*; Perun, T. J., Propst, C. L., Eds.; Marcel Dekker, Inc.: New York, 1989; p 55.
(178) Hopfinger, A. J.; Pearlstein, R. A. *J. Comput. Chem.* **1984**, *5*, 486.
(179) Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; Case, D. A. *J. Comput. Chem.* **1986**, *7*, 230.
(180) Hall, A.; Pavitt, N. *J. Comput. Chem.* **1984**, *5*, 411.
(181) Jorgensen, W. L.; Swenson, C. J. *J. Am. Chem. Soc.* **1985**, *107*, 1489.
(182) Jorgensen, W. L.; Madura, J. D.; Swenson, C. J. *J. Am. Chem. Soc.* **1984**, *106*, 6638.
(183) Jorgensen, W. L.; Swenson, C. J. *J. Am. Chem. Soc.* **1985**, *107*, 569.
(184) Boyd, D. B.; Lipkowitz, K. B. *J. Chem. Educ.* **1982**, *59*, 269.
(185) Hagler, A. T.; Maple, J. R.; Thacher, T. S.; Fitzgerald, G. B.; Dinur, U. *Computer Simulation of Biomolecular Systems: Theoretical and Experimental Applications*; van Gunsteren, W. F., Weiner, P. K., Eds.; ESCOM Science Publishers B. V.: Leiden, 1989; p 149.
(186) Buckert, U.; Allinger, N. L. *Molecular Mechanics*; American Chemical Society: Washington, DC, 1982.
(187) Sadek, M.; Munro, S. *J. Comput.-Aided Mol. Des.* **1988**, *2*, 81.
(188) Merz, K.; Kollman, P. A. *J. Am. Chem. Soc.* **1989**, in press.
(189) van Gunsteren, W. F. *Computer Simulation of Biomolecular Systems: Theoretical and Experimental Applications*; van Gunsteren, W. F.; Weiner, P. K., Eds.; ESCOM Science Publishers B. V.: Leiden, 1989; p 27.
(190) Pearlman, D. A.; Kollman, P. A. ref 189, p 101.